

Série A, n° 2618

N° d'Ordre : 3490

THÈSES

présentées à la

FACULTÉ DES SCIENCES DE L'UNIVERSITÉ DE PARIS

pour obtenir

le Grade de DOCTEUR ès SCIENCES MATHÉMATIQUES

par

Henri MARCHAND

Ancien élève de l'École Polytechnique

Agrégé de l'Université

1^{re} THÈSE : Etude mathématique de l'évolution d'une
population mendélienne soumise à une
loi d'union sélective.

2^{me} THÈSE : Propositions données par la Faculté.

Soutenues le 4 Novembre 1953 devant la Commission d'examen

MM. J. PERES Président

G. DARMOIS . . }
R. FORTET . . . } Examineurs
G. MALECOT . . }

Série A, n° 2618
N° d'Ordre : 3490

THÈSES

présentées à la

FACULTÉ DES SCIENCES DE L'UNIVERSITÉ DE PARIS

pour obtenir

le Grade de DOCTEUR ès SCIENCES MATHÉMATIQUES

par

Henri MARCHAND

Ancien élève de l'École Polytechnique

Agrégé de l'Université

1^{re} THÈSE : Etude mathématique de l'évolution d'une
population mendélienne soumise à une
loi d'union sélective.

2^{me} THÈSE : Propositions données par la Faculté.

Soutenues le 4 Novembre 1953 devant la Commission d'examen

MM. J. PERES Président

G. DARMOIS . .	} Examineurs
R. FORTET . . .	
G. MALECOT . .	



FACULTÉ DES SCIENCES DE L'UNIVERSITÉ DE PARIS

Doyen. M. A. CHATELET.

PROFESSEURS

G. JULIA.	T Analyse supérieure et Algèbre supérieure.	CHATELET	T Arithmétique et théorie des nombres.
A. DENJOY	T Théorie des Fonctions.	EPHRUSSI.	T Génétique.
L. LUTAUD	T Géographie physique et Géologie dynamique.	WURMSER	T Biologie physico-chimique.
E. DARMOIS.	T Enseignement de Physique.	RIVIÈRE	Géologie (P.C.B.).
Robert LÉVY	T Physiologie comparée.	GAUTHERET.	T Biologie végétale (P.C.B.).
FOCH	T Mécanique physique et expérimentale.	LUCAS	T Physique.
PAUTHENIER.	T Electrotechnique générale.	A. THOMAS.	T Evolution des êtres organisés.
DE BROGLIE.	T Théories physiques.	ARNULF	Optique appliquée.
JOB	T Chimie générale.	Max MORAND.	Physique.
PRENANT.	T Anatomie et Histologie comparées.	SOLEILLET	T Physique.
GARNIER.	T Géométrie supérieure.	FORTIER	T Mécanique expérimentale des fluides.
PÈRES	T Mécanique des fluides et applications.	DANJON	T Astronomie.
LAUGIER	T Physiologie générale.	FROMAGEOT.	T Chimie biologique.
M. CURIE.	T Physique (P. C. B.).	LAPORTE.	T Physique générale et Radio-activité.
G. RIBAUD	T Hautes températures.	JANET	T Mathématiques générales.
CROZE	T Physique théorique et physique céleste.	PETIT	T Biologie maritime.
DUPONT	T Théories chimiques.	QUENEY	T Météorologie et dynamique atmosphérique.
VALIRON.	T Calcul différentiel et Calcul intégral.	GALLIEN.	T Biologie animale (P.C.B.).
BARRABÉ.	T Géologie structurale et géologie appliquée.	EICHHORN	Biologie végétale (P.C.B.).
G. DARMOIS	T Calcul des probabilités et Physique mathématique.	DE CUGNAC.	Biologie végétale (P.C.B.).
Jacques BOURCART. T	Géographie physique et Géologie dynamique.	M ^{lle} CAUCHOIS	T Chimie-Physique.
AUBEL.	T Chimie biologique.	THELLIER.	Physique du Globe.
M ^{me} JOLIOT-CURIE T	Physique générale et Radio-activité.	AUDUBERT	Electrochimie.
PLANTEFOL	T Botanique.	L'HÉRITIER.	Génétique.
CABANNES	T Recherches physiques.	GRIVET.	T Radio-Electricité.
GRASSÉ.	T Evolution des êtres organisés.	PONCIN.	T Mécanique des fluides.
PRÉVOST	T Chimie organique.	THIRY.	T Mécanique appliquée.
BOULIGAND.	T Application de l'analyse à la Géométrie.	DUBREIL	T Mathématiques générales.
CHAUDRON.	T Chimie appliquée.	QUELET	T Chimie (P.C.B.).
WYART.	T Minéralogie.	CAGNIARD	Géophysique appliquée.
TEISSIER	T Zoologie.	CHAMPETIER	Chimie appliquée.
MANGENOT.	T Biologie végétale.	CUVILLIER	Géologie structurale et Géologie appliquée.
P. AUGER.	T Physique quantique et relativité.	JUNG	T Pétrographie.
MONNIER.	T Physiologie générale.	TRILLAT.	Physique (P.C.B.).
PIVETEAU.	T Géologie.	WIEMANN	T Chimie (P.C.B.).
ROCARD	T Physique.	JACQUINOT.	T Physique (P.C.B.).
H. CARTAN.	T Mathématiques (E.N. Supérieure).	VASSY	Physique de l'atmosphère.
LAFFITTE.	T Chimie générale.	DESTOUCHES	Théories physiques.
FAVARD	T Mécanique générale.	M. PRUVOST	T Géologie.
COULOMB.	T Physique du Globe.	AMIEL.	Chimie (P.C.B.).
M ^{lle} COUSIN	T Biologie animale (P.C.B.).	HOCART	T Minéralogie.
CHRÉTIEN.	T Chimie minérale.	J.-P. MATHIEU	Physique (P.C.B.).
P. DRACH.	T Zoologie.	COUTEAUX	Biologie animale (P.C.B.).
KASTLER	T Physique.	MAY.	Biologie animale (P.C.B.).
		CHOQUET.	Calcul différentiel.
		FELDMANN.	Biologie (P.C.B.).
		GUINIER.	Physique (P.C.B.).
		JOST.	Biologie (P.C.B.).
		FORTET	Calcul des probabilités.
		SCHWARTZ	Méthodes mathématiques de la Physique.

Secrétaire général Ch. MONIER.

A MONSIEUR GEORGES DARMOIS

Professeur à la Sorbonne

A MONSIEUR GUSTAVE MALÉCOT

Professeur à la Faculté des Sciences de Lyon

A MA MÈRE

A MA FEMME

PREMIÈRE THÈSE

ÉTUDE MATHÉMATIQUE DE L'ÉVOLUTION D'UNE POPULATION MENDELIIENNE SOUMISE A UNE LOI D'UNION SÉLECTIVE

AVANT-PROPOS

Au cours de ses travaux sur la génétique des populations, M. Sewall WRIGHT a défini et étudié le niveau d'adaptation (« adaptive value ») d'une population, concernant un caractère « primaire » déterminé ⁽¹⁾ ; il a d'autre part traité le problème de l'évolution mendélienne dans l'hypothèse où les croisements ont lieu au hasard (« random mating ») et où la sélection est fonction du niveau d'adaptation ⁽²⁾.

Nous nous proposons ici d'étudier l'évolution d'une population mendélienne non panmictique, mais où la loi d'union des individus dépend elle-même du niveau d'adaptation de ceux-ci.

Nous supposerons connues les notions de gènes, loci, facteurs mendéliens, ainsi que les lois de Mendel. Nous comparerons fréquemment nos conclusions aux résultats établis par M. MALÉCOT dans ses travaux sur la théorie mathématique de l'hérédité, notamment dans l'hypothèse du random mating.

Dans la première partie, nous aborderons le problème d'un point de vue déterministe, en supposant la population très nombreuse.

Après avoir, au chapitre premier, défini la loi d'union adoptée et établi ses propriétés générales, nous traiterons aux chapitres II et III de l'évolution d'une population lorsque le caractère primaire ne dépend que d'un facteur, en essayant de dégager le rôle joué par le degré de dominance de l'un des gènes allèles, puis nous verrons comment les résultats obtenus sont modifiés si l'on admet l'existence de mutations.

(1) Sewall WRIGHT, « The analysis of variance and correlations between relatives with respect to deviations from an optimum », *Journal of Genetics*, vol. XXX, 1935, p. 243 à 256.

(2) *Id.*, « Evolution in populations in approximate equilibrium », *Journal of Genetics*, vol. XXX, 1935, p. 257 à 266.

Dans les chapitres IV et V nous établirons les valeurs asymptotiques des probabilités d'association des gènes lorsque le caractère primaire dépend de plusieurs facteurs et nous en déduirons les valeurs des corrélations usuelles dans une population en équilibre.

Enfin, au chapitre VI, nous comparerons notre hypothèse de loi d'union sélective à l'hypothèse de M. Sewall WRIGHT.

Dans une deuxième partie, nous reprendrons la question d'un point de vue stochastique en étudiant l'effet de la loi d'union considérée sur l'évolution d'une population d'effectif limité constant.

Nous établirons au chapitre premier l'équation qui détermine la loi de probabilité de la fréquence des gènes à un instant donné, en supposant que le caractère primaire ne dépend que d'un facteur. Nous montrerons l'existence d'une loi de probabilité asymptotique et nous indiquerons sa forme.

Aux chapitres II et III nous déterminerons la loi asymptotique en faisant diverses hypothèses sur le niveau d'adaptation de la population.

Enfin, nous verrons au chapitre V comment les résultats sont transformés par la présence soit d'un certain degré de dominance, soit d'un phénomène de mutations.

* *

Je tiens à exprimer ici ma reconnaissance à M. G. DARMOIS qui, après m'avoir montré l'intérêt qu'offre ce domaine de recherches, n'a cessé de m'encourager dans mon travail ; je désire témoigner ma très vive gratitude à M. G. MALÉCOT, qui m'a constamment dirigé dans mon étude et dont les conseils m'ont été si précieux ; j'adresse mes sincères remerciements à M. E. BOREL qui a présenté à l'Académie des Sciences les principaux résultats⁽¹⁾, à M. PÉRÈS qui a bien voulu présider le jury, et à M. FORTET qui a accepté de se joindre à MM. PÉRÈS, G. DARMOIS et G. MALÉCOT pour constituer le jury.

(1) *C.R. Ac. Sc. Paris* : 231, p. 1029 et 1210 ; 233, p. 1259 et 1356 ; 235, p. 863.

PREMIERE PARTIE

POPULATION TRÈS NOMBREUSE

CHAPITRE PREMIER

DÉFINITION D'UNE LOI D'UNION SÉLECTIVE

Considérons une population très nombreuse. Soit H un caractère mesurable. Désignons par π_i la fréquence des individus de la population chez lesquels ce caractère a la valeur *centrée* y_i ; la fréquence des couples formés d'un individu de valeur y_1 et d'un individu de valeur y_2 serait dans le cas du random mating $\pi_1 \cdot \pi_2$.

Nous nous proposons d'étudier une population dans laquelle la fréquence de tels couples est :

$$(1) \quad \pi_1 \cdot \pi_2 (1 + \mu y_1 + \mu y_2)$$

μ étant un coefficient constant petit.

Remarquons que la somme des fréquences de tous les couples est égale à 1 ; en effet :

$$\sum_1 \sum_2 \pi_1 \pi_2 (1 + \mu y_1 + \mu y_2) = \sum_1 \pi_1 \sum_2 \pi_2 + \mu \sum_2 \pi_2 \sum_1 \pi_1 y_1 + \mu \sum_1 \pi_1 \sum_2 \pi_2 y_2 = 1$$

puisque :

$$\sum_1 \pi_1 = \sum_2 \pi_2 = 1 ; \quad \sum_1 \pi_1 y_1 = \sum_2 \pi_2 y_2 = 0$$

y_1 et y_2 étant les valeurs centrées du caractère H .

* * *

Considérons alors deux catégories quelconques d'individus de la population ; désignons les fréquences totales respectives des individus de chacune de ces catégories par :

$$\pi' = \sum' \pi_1 ; \quad \pi'' = \sum'' \pi_2$$

Les valeurs moyennes du caractère H sont dans chaque catégorie :

$$\bar{y}' = \frac{\sum' \pi_1 y_1}{\sum' \pi_1} ; \quad \bar{y}'' = \frac{\sum'' \pi_2 y_2}{\sum'' \pi_2}$$

La fréquence totale des unions entre individus de ces deux catégories sera, en groupant dans chaque catégorie les individus chez lesquels le caractère H a la même valeur :

$$\begin{aligned}\Sigma' \Sigma'' \pi_1 \pi_2 (1 + \mu y_1 + \mu y_2) &= \Sigma' \pi_1 \Sigma'' \pi_2 + \mu \Sigma'' \pi_2 \Sigma' \pi_1 y_1 + \mu \Sigma' \pi_1 \Sigma'' \pi_2 y_2 \\ &= \pi' \pi'' + \mu \pi'' \pi' y' + \mu \pi' \pi'' y''\end{aligned}$$

et l'on voit finalement que la fréquence des accouplements entre deux catégories quelconques d'individus est donnée par l'expression :

$$(2) \quad \pi' \cdot \pi'' (1 + \mu y' + \mu y'')$$

Nous représenterons l'expression précédente (2) par le produit symbolique :

$$\pi' * \pi''$$

Ce produit symbolique, évidemment commutatif, est également distributif ; en effet, supposons que la catégorie Σ'' soit la somme de deux catégories Σ''' et Σ^{IV} ; on peut écrire :

$$\begin{aligned}\pi'' &= \pi''' + \pi^{IV} ; & \pi''' &= \Sigma''' \pi_2 ; & \pi^{IV} &= \Sigma^{IV} \pi_2 \\ \pi'' y'' &= \Sigma'' \pi_2 y_2 = \Sigma''' \pi_2 y_2 + \Sigma^{IV} \pi_2 y_2\end{aligned}$$

et comme on a :

$$\begin{aligned}y''' &= \frac{\Sigma''' \pi_2 y_2}{\pi'''} ; & y^{IV} &= \frac{\Sigma^{IV} \pi_2 y_2}{\pi^{IV}} \\ \pi'' y'' &= \pi''' y''' + \pi^{IV} y^{IV}\end{aligned}$$

donc :

$$\begin{aligned}\pi' * \pi'' &= \pi' \pi'' + \mu (\pi' \pi'' y'' + \pi' \pi'' y') \\ &= \pi' (\pi''' + \pi^{IV}) + \mu \pi' (\pi''' y''' + \pi^{IV} y^{IV}) + \mu (\pi''' + \pi^{IV}) \pi' y' \\ &= \pi' \pi''' (1 + \mu y' + \mu y''') + \pi' \pi^{IV} (1 + \mu y' + \mu y^{IV}) \\ &= \pi' * \pi''' + \pi' * \pi^{IV}\end{aligned}$$

Remarquons d'autre part que la loi d'union adoptée est *sélective* ; en effet, la fréquence d'accouplement entre un individu de valeur déterminée y_1 et un individu quelconque de la population ($\pi'' = 1$; $y'' = 0$) est :

$$\pi_1 (1 + \mu y_1)$$

Elle n'est pas égale à π_1 comme cela se produirait s'il n'y avait pas de sélection : *tous les individus n'ont donc pas la même probabilité de trouver des conjoints.*

* * *

Supposons maintenant que nous classions les individus de la population suivant les trois phases d'un facteur déterminé. Désignons par A et a les deux états possibles des gènes correspondants. Soit P, 2Q, R les fréquences respectives des trois génotypes AA, Aa, aa

dans une génération déterminée, I, J, K les valeurs moyennes centrées du caractère H dans ces trois catégories. On a les relations :

$$PI + 2 QJ + RK = 0 ; \quad P + 2 Q + R = 1$$

Posons d'autre part :

$$p = P + Q ; \quad q = Q + R \quad (p + q = 1)$$

p et q représentent les fréquences des gènes A et a.

Enfin, soit T et S les valeurs moyennes centrées du caractère H dans deux catégories contenant l'une les homozygotes AA et la moitié des hétérozygotes, l'autre les homozygotes aa et la moitié des hétérozygotes. T et S ont pour expressions :

$$T = \frac{PI + QJ}{p} ; \quad S = \frac{QJ + RK}{q} \quad (pT + qS = 0)$$

Nous représenterons par des lettres accentuées P', Q', R'... les quantités analogues concernant la génération suivante.

Si nous supposons que tous les couples ont la même fécondité, la fréquence P' est égale à la somme des produits des fréquences des accouplements par la probabilité qu'a chaque enfant résultant de cet accouplement d'être un homozygote AA. Il en est de même en ce qui concerne les fréquences Q' et R'. Nous avons donc :

$$P' = P * P + 1/2 P * 2 Q + 1/2 2 Q * P + 1/4 2 Q * 2 Q$$

et les deux formules analogues :

$$2 Q' = 1/2 P * 2 Q + 1/2 2 Q * P + P * R + R * P + 1/2 2 Q * 2 Q + 1/2 2 Q * R + 1/2 R * 2 Q$$

$$R' = R * R + 1/2 2 Q * R + 1/2 R * 2 Q + 1/4 2 Q * 2 Q$$

D'où en tenant compte des propriétés du produit symbolique :

$$\begin{cases} P' = P * P + 2 P * Q + Q * Q = (P + Q) * (P + Q) = p * p \\ 2 Q' = 2 Q * R + 2 P * Q + 2 P * R + 2 Q * Q = 2 (P + Q) * (Q + R) = 2 p * q \\ R' = R * R + 2 Q * R + Q * Q = (Q + R) * (Q + R) = q * q \end{cases}$$

Or, T et S sont précisément les valeurs moyennes centrées du caractère H dans les deux catégories de fréquences respectives $P + Q = p$ et $Q + R = q$; nous obtenons donc, en explicitant les produits symboliques, les expressions suivantes :

(3)

$$\begin{aligned} P' &= p^2 (1 + 2 \mu T) \\ 2 Q' &= 2 p q (1 + \mu T + \mu S) \\ R' &= q^2 (1 + 2 \mu S) \end{aligned}$$

On en déduit immédiatement les expressions de p' et q' :

(4)

$$\begin{aligned} p' &= P' + Q' = p (1 + \mu T) \\ q' &= Q' + R' = q (1 + \mu S) \end{aligned}$$

en tenant compte des relations $p + q = 1$ et $pT + qS = 0$.

La variation Δp de la fréquence p du gène A d'une génération à la suivante est donc :

(5)

$$\Delta p = p' - p = \mu p T$$

*
* *

Nous supposons désormais le coefficient μ suffisamment petit pour que l'on puisse négliger son carré. Les relations (4) donnent alors :

$$\begin{cases} p'^2 = p^2 (1 + 2 \mu T) + 0 (\mu^2) \\ 2 p' q' = 2 p q (1 + \mu T + \mu S) + 0 (\mu^2) \\ q'^2 = q^2 (1 + 2 \mu S) + 0 (\mu^2) \end{cases}$$

D'où, en tenant compte des relations (3) :

$$\begin{cases} p'^2 = P' + 0 (\mu^2) \\ 2 p' q' = 2 Q' + 0 (\mu^2) \\ q'^2 = R' + 0 (\mu^2) \end{cases}$$

Avec l'approximation adoptée, on voit que la loi de Hardy est vérifiée :

(6)

$$P = p^2 ; \quad 2 Q = 2 p q ; \quad R = q^2$$

*
* *

Population à sexes séparés. — Nous avons supposé dans ce qui précède que chaque individu présente indifféremment les deux sexes ; les résultats trouvés s'étendent au cas où les individus sont à sexes séparés.

Nous supposons que les caractères envisagés ne sont pas liés au sexe. Nous affecterons de l'indice antérieur 1 les quantités qui se rapportent aux mâles, de l'indice antérieur 2 celles qui sont relatives aux femelles. Les quantités sans indice concernent l'ensemble de la population sans considération de sexe.

Considérons alors la loi d'union de la forme :

$$1^{\pi} * 2^{\pi} (1 + \mu 1y + \mu 2y)$$

où ${}_1y$ et ${}_2y$ sont les moyennes centrées du caractère H, ces moyennes étant rapportées à la moyenne générale du caractère pour chacun des sexes, c'est-à-dire que l'on a :

$$\Sigma {}_1\pi {}_1y = \Sigma {}_2\pi {}_2y = 0$$

On peut alors définir comme plus haut le produit symbolique ${}_1\pi' * {}_2\pi''$ et établir qu'il est distributif.

D'autre part, on a les relations :

$${}_1P {}_1I + {}_2Q {}_1J + {}_1R {}_1K = 0 ; \quad {}_1P + {}_2Q + {}_1R = 1 ; \quad {}_1p {}_1T + {}_1q {}_1S = 0$$

et les relations analogues pour la population femelle. On en déduit, en classant alors les mâles et les femelles suivant leur génotype, les formules semblables aux formules (3) :

$$(3') \quad P' = {}_1p {}_2p (1 + \mu {}_1T + \mu {}_2T) ; \quad 2Q' = {}_1p {}_2q (1 + \mu {}_1T + \mu {}_2S) + {}_2p {}_2q (1 + \mu {}_1S + \mu {}_2T)$$

et finalement :

$$(4') \quad \begin{aligned} p' &= 1/2 [{}_1p (1 + \mu {}_1T) + {}_2p (1 + \mu {}_2T)] \\ q' &= 1/2 [{}_1q (1 + \mu {}_1S) + {}_2q (1 + \mu {}_2S)] \end{aligned}$$

Mais, si l'on remarque que la population étant très nombreuse, les fréquences des gènes et des génotypes dans une génération sont sensiblement égales aux probabilités de ceux-ci, et sont, par conséquent, les mêmes pour la population mâle et la population femelle, on a, à partir de la deuxième génération :

$${}_1p = {}_2p = p ; \quad {}_1q = {}_2q = q ; \quad {}_1P = {}_2P = P ; \quad {}_1T = {}_2T = T ; \dots$$

et l'on retrouve les expressions (5) et (6).

Nous supposons donc dans la suite de cette première partie (population nombreuse), la population à sexes réunis.

*
* *

Précisons maintenant la nature du caractère H envisagé. Soit (S) un caractère mesurable que nous appellerons « caractère primaire » ; H sera le caractère « secondaire » mesuré par le carré changé de signe de l'écart entre (S) et une valeur optimum que nous désignerons par $\underline{O}$:

$$H = - [(S) - \underline{O}]^2$$

En écrivant M (S) et M (H) les moyennes de (S) et de H, les valeurs centrées de H sont

$$y_i = - [(S_i) - \underline{O}]^2 - M (H)$$

Le coefficient μ qui caractérise l'union sélective est > 0 car nous supposons que l'individu éloigné de l'optimum est défavorisé.

CHAPITRE II

ÉVOLUTION D'UNE POPULATION NOMBREUSE ROLE DE LA DOMINANCE

(Le caractère primaire ne dépendant que d'un facteur)

Nous supposerons dans ce chapitre et dans le suivant que le caractère primaire (S) ne dépend que d'un facteur.

Classons encore les individus suivant les trois génotypes AA, Aa, aa et soit 2α , $\alpha(1+d)$, 0, les valeurs du caractère (S) correspondant à chaque génotype ; le coefficient d caractérise le degré de dominance du gène A. On a :

$$0 \leq d \leq 1$$

Les valeurs des caractères primaire et secondaire suivant le génotype de l'individu sont données par le tableau suivant :

	AA	Aa	aa
(S)	2α	$\alpha(1+d)$	0
H	$-(2\alpha - \underline{O})^2$	$-\alpha(1+d - \underline{O})^2$	$-\underline{O}^2$

D'où, en donnant aux fréquences des génotypes les valeurs approchées établies au chapitre précédent (I - 6) :

$$(1) \quad M(H) = -2\alpha^2p[2p + (1+d)^2q] + 4\alpha\underline{O}p(1+dq) - \underline{O}^2$$

Les valeurs centrées du caractère H sont alors :

$$I = -(2\alpha - \underline{O})^2 - M(H); \quad J = -\alpha(1+d - \underline{O})^2 - M(H); \quad K = -\underline{O}^2 - M(H)$$

et nous pouvons calculer T :

$$T = (PI + QJ)/p = pI + qJ = -(2\alpha - \underline{O})^2p - [\alpha(1+d - \underline{O})^2q - M(H)]$$

ce qui, après calculs et simplifications, conduit à l'expression :

$$(2) \quad T = 2\alpha\underline{O}q(1+d - 2dp) - \alpha^2q[4p + (1+d)^2(1-2p)]$$

L'équation (I - 5) donne alors :

$$\Delta p = \mu p T = \mu \alpha p (1 - p) \left\{ p [-4 d \underline{O} - 4 \alpha + 2 \alpha (1 + d)^2] + 2 (1 + d) \underline{O} - \alpha (1 + d)^2 \right\}$$

Nous conviendrons de poser :

$$(3) \quad \begin{cases} h = \underline{O}/\alpha - 1 \\ \beta = 1/2 - h/(d^2 - 2 dh - 1) \\ k = -2 \mu \alpha^2 (d^2 - 2 dh - 1) \end{cases}$$

L'expression de Δp devient alors, si l'on suppose $h \neq (d^2 - 1)/2 d$:

$$(4) \quad \Delta p = k p (p - 1) (p - \beta)$$

C'est là l'expression de la variation de la probabilité du gène A, et donc de sa fréquence puisque la population est très nombreuse, lorsqu'on passe d'une génération à la suivante.

*
* *
*

Proposons-nous maintenant d'étudier la composition asymptotique de la population. Nous distinguerons plusieurs cas suivant les valeurs des paramètres h, β, k . Nous supposons dans cette discussion $0 < d < 1$, nous réservant de traiter ensuite le cas de l'absence de dominance ($d = 0$) et celui de la dominance complète ($d = 1$).

Les valeurs stationnaires possibles de p sont :

$$p = 0; \quad p = 1; \quad p = \beta$$

1^{er} cas : $k < 0$.

Ce cas se produit si $d^2 - 2 dh - 1 > 0$, c'est-à-dire si :

$$(5) \quad h < (d^2 - 1)/2 d$$

Δp est alors du signe de $p - \beta$. Or, dans le cas envisagé, β est toujours supérieur à 1. En effet, l'inégalité $\beta > 1$ s'écrit :

$$2 h < 1 + 2 dh - d^2$$

puisque $d^2 - 2 dh - 1$ est > 0 . D'où :

$$2 h (1 - d) < 1 - d^2$$

et :

$$h < (1 + d)/2$$

Or, cette inégalité est une conséquence de (5), car :

$$(d^2 - 1)/2 d < (1 + d)/2 \quad \text{puisque } d < 1.$$

Donc $p - \beta$ est toujours < 0 et Δp est < 0 : p décroît constamment.

Soit alors p_0 la fréquence du gène A dans la génération initiale ; l'expression :

$$-\Delta p/p = k(1-p)(p-\beta)$$

positive dans l'intervalle $(0, p_0)$, présente dans cette intervalle une valeur minimum positive m [égale à $k(1-p_0)(p_0-\beta)$]. Désignons par p_{n-1} et p_n les valeurs de p dans les générations de rang $n-1$ et n ; on a :

$$(p_{n-1} - p_n)/p_{n-1} > m$$

d'où :

$$p_n < (1-m)p_{n-1} < (1-m)^2 p_{n-2} < \dots < p_0(1-m)^n$$

Or, m est de l'ordre de k , c'est-à-dire de μ qui est supposé petit ; on a donc $m < 1$. Par conséquent p tend vers 0 plus rapidement que $(1-m)^n$ lorsque n augmente indéfiniment.

Le gène A est éliminé quel que soit p_0 (1).

2^e cas : $k > 0$.

Ce cas se produit si :

$$(6) \quad h > (d^2 - 1)/2d$$

Le signe de Δp est alors celui de $(\beta - p)$.

a) $\beta < 0$,

c'est-à-dire $h/(d^2 - 2dh - 1) > 1/2$

ou :

$$h < (d - 1)/2$$

ce qui est compatible avec (6), car :

$$(d^2 - 1)/2d < (d - 1)/2 \quad \text{puisque} \quad d < 1.$$

Cette condition exprime que le génotype aa est le mieux adapté.

Δp est alors constamment < 0 . Le raisonnement précédent montre encore que p tend vers 0 plus rapidement que $(1-m)^n$ lorsque n augmente indéfiniment, m étant le plus petit des deux nombres > 0 : $-k\beta$ et $k(1-p_0)(p_0-\beta)$.

Il y a donc fixation asymptotique du gène a.

(1) Remarquons que l'on a : $h + 1 = \frac{O}{\alpha} < \frac{d^2 - 1}{2d} + 1 < \frac{d + 1}{2}$ = ce qui prouve que dans ce cas le génotype aa est le mieux adapté.

b) $\beta > 1$,

c'est-à-dire $1/2 - h/(d^2 - 2dh - 1) > 1$

ou :

$$h > (d + 1)/2$$

inégalité qui entraîne l'inégalité (6) et qui exprime que le génotype AA est le mieux adapté.

La variation Δp est alors > 0 . En posant $q = 1 - p$, on obtient :

$$\Delta q = -\Delta p = -k(1 - q)(-q)(1 - q - \beta) = kq(q - 1)(q - 1 + \beta)$$

Or, $(1 - \beta)$ est < 0 . Nous sommes ramenés au cas précédent en remplaçant p par q et β par $(1 - \beta)$. Donc q tend vers 0 plus rapidement que $(1 - m)^n$ quand n augmente indéfiniment, m étant le plus petit des deux nombres > 0 : $-k(1 - \beta)$ et $k p_0 (\beta - p_0)$.

Il y a donc *fixation asymptotique du gène A*.

c) $0 < \beta < 1$,

c'est-à-dire :

$$(d - 1)/2 < h < (d + 1)/2$$

inégalités qui entraînent (6) et expriment que l'hétérozygote Aa est le mieux adapté des trois génotypes.

Or $\Delta p = \Delta(p - \beta)$. Nous pouvons donc écrire :

$$-\frac{\Delta(p - \beta)}{p - \beta} = kp(1 - p) > 0$$

L'expression précédente dans l'intervalle (p_0, β) a une valeur minimum positive m [égale soit à $k p_0 (1 - p_0)$, soit à $k \beta (1 - \beta)$].

Posons :

$$p_n - \beta = z_n$$

On a, quel que soit n :

$$(z_{n-1} - z_n)/z_{n-1} > m$$

d'où :

$$|z_n| < (1 - m) |z_{n-1}| < (1 - m)^2 |z_{n-2}| < \dots < (1 - m)^n |z_0|$$

ce qui montre que z tend vers 0 plus rapidement que $(1 - m)^n$ lorsque n augmente indéfiniment

Donc p a pour limite β . La population tend vers un équilibre asymptotique dans lequel les deux gènes coexistent.

d) Cas limite $\beta = 0$,

c'est-à-dire :

$$h = (d - 1)/2$$

Les génotypes Aa et aa sont alors équidistants de 0.

L'expression de la variation Δp devient :

$$\Delta p = kp^2 (p - 1)$$

avec : $k = -2\mu\alpha^2 (d - 1) > 0$

Donc Δp est constamment < 0 ; la fréquence p tend vers 0 : il y a *fixation asymptotique du gène a*.

e) Cas limite $\beta = 1$,

c'est-à-dire :

$$h = (d + 1)/2$$

Les génotypes AA et Aa sont alors équidistants de 0.

L'expression de la variation Δp devient :

$$\Delta p = kp (p - 1)^2$$

avec : $k = 2\mu\alpha^2 (d + 1) > 0$

Donc Δp est constamment > 0 ; la fréquence p tend vers 1 : il y a *fixation asymptotique du gène A*.

3^e cas : $h = (d^2 - 1)/2d$.

Dans ce cas, les formules (3) ne peuvent être considérées, car le paramètre β n'existe pas.

Le calcul de la variation Δp conduit au résultat suivant :

$$\Delta p = \mu p T = -2\mu\alpha^2 h p (p - 1)$$

en remplaçant dans l'expression (2) de T la valeur optimum 0 par :

$$\propto \frac{(d + 1)^2 - 2}{2d}$$

Δp est donc < 0 quel que soit p . La fréquence p tend vers 0, et il y a *fixation asymptotique du gène a*.

* *

Cette discussion montre, en résumé, qu'il y a fixation du gène A (ou du gène a) lorsque le génotype le mieux adapté est l'homozygote AA (ou aa) ; et qu'il y a équilibre asymptotique avec coexistence des deux gènes, si le génotype hétérozygote Aa est le mieux adapté.

* *

Cas où il n'y a pas de dominance : $d = 0$.

Les formules (2) et (3) deviennent :

$$(2') \quad T = 2 \frac{O}{\alpha} q - \alpha^2 q (1 + 2p)$$

$$(3') \quad \begin{aligned} h &= \frac{O}{\alpha} - 1 \\ \beta &= h + 1/2 = \frac{O}{\alpha} - 1/2 \\ k &= 2 \mu \alpha^2 \end{aligned}$$

On a donc toujours $k > 0$; Δp a le signe de $(\beta - p)$. La discussion précédente (2^e cas, $k > 0$) donne alors les résultats suivants :

Si $\frac{O}{\alpha} \leq \alpha/2$, il y a fixation asymptotique du gène a ;

Si $\frac{O}{\alpha} \geq 3\alpha/2$, il y a fixation asymptotique du gène A ;

Si $\alpha/2 < \frac{O}{\alpha} < 3\alpha/2$, il y a un équilibre asymptotique avec coexistence des deux allèles, la fréquence limite du gène A étant $\frac{O}{\alpha} - 1/2$.

* *

Rapidité d'évolution de la population.

Afin de préciser davantage la rapidité avec laquelle la population atteint son état d'équilibre, remarquons que Δp est la différence première de p , fonction du temps, lorsqu'on prend comme unité de temps la durée d'une génération ; l'intégration de cette différence première se ramène sensiblement à l'intégration de :

$$\frac{dp}{dt} = kp(p-1)(p-\beta)$$

c'est-à-dire à la quadrature :

$$kt = \int_{p_0}^p \frac{dp}{p(p-1)(p-\beta)}$$

ce qui donne :

$$(7) \quad kt = \frac{1}{\beta} L \frac{p}{p_0} + \frac{1}{1-\beta} L \frac{1-p}{1-p_0} + \frac{1}{\beta(\beta-1)} L \frac{p-\beta}{p_0-\beta}$$

Nous distinguerons plusieurs cas suivant la valeur asymptotique de la fréquence p :

1° $k < 0$ et $\beta > 1$ ou $k > 0$ et $\beta < 0$: p tend vers 0.

L'expression (7) montre alors que :

$$L \frac{p}{p_0} = \beta kt - \frac{\beta}{1-\beta} L \frac{1-p}{1-p_0} + \frac{1}{1-\beta} L \frac{p-\beta}{p_0-\beta}$$

ou :

$$\frac{p}{p_0} = e^{\beta kt} \cdot \left(\frac{1-p}{1-p_0} \right)^{\frac{\beta}{\beta-1}} \cdot \left(\frac{p-\beta}{p_0-\beta} \right)^{\frac{1}{1-\beta}}$$

On a donc, lorsque t augmente indéfiniment :

$$p \sim C e^{\beta kt}$$

en posant :

$$C = p_0 \left[(1-p_0)^{-\beta} \cdot \frac{p_0-\beta}{-\beta} \right]^{\frac{1}{\beta-1}}$$

Comme βk est < 0 , p décroît et tend asymptotiquement vers 0 comme la fonction exponentielle décroissante :

$$e^{\beta kt}$$

2° $k > 0$ et $\beta > 1$: p tend vers 1.

L'expression (7) montre alors que :

$$L \frac{1-p}{1-p_0} = (1-\beta) kt - \frac{1-\beta}{\beta} L \frac{p}{p_0} + \frac{1}{\beta} L \frac{p-\beta}{p_0-\beta}$$

ou :

$$\frac{1-p}{1-p_0} = e^{(1-\beta) kt} \cdot \left(\frac{p}{p_0} \right)^{\frac{\beta-1}{\beta}} \cdot \left(\frac{p-\beta}{p_0-\beta} \right)^{\frac{1}{\beta}}$$

On a donc, lorsque t augmente indéfiniment :

$$1 - p \sim C e^{(1 - \beta) kt}$$

en posant :

$$C = (1 - p_0) \left[\frac{1 - \beta}{\beta - 1} \frac{1}{(p_0 - \beta)} \right]^{\frac{1}{\beta}}$$

Comme $(1 - \beta) k$ est < 0 , $(1 - p)$ décroît et tend asymptotiquement vers 0 comme la fonction exponentielle décroissante :

$$e^{(1 - \beta) kt}$$

3° $k > 0$ et $0 < \beta < 1$: p tend vers β .

L'expression (7) donne alors :

$$L \frac{p - \beta}{p_0 - \beta} = \beta (\beta - 1) kt - (\beta - 1) L \frac{p}{p_0} + \beta L \frac{1 - p}{1 - p_0}$$

ou :

$$\frac{p - \beta}{p_0 - \beta} = e^{\beta (\beta - 1) kt} \cdot \left(\frac{p}{p_0} \right)^{1 - \beta} \cdot \left(\frac{1 - p}{1 - p_0} \right)^{\beta}$$

On a donc, lorsque t augmente indéfiniment :

$$|p - \beta| \sim C e^{\beta (\beta - 1) kt}$$

en posant :

$$C = |p_0 - \beta| \left(\frac{\beta}{p_0} \right)^{1 - \beta} \left(\frac{1 - \beta}{1 - p_0} \right)^{\beta}$$

Comme $k\beta(\beta - 1)$ est < 0 , $|p - \beta|$ décroît et tend asymptotiquement vers 0 comme la fonction exponentielle décroissante :

$$e^{\beta (\beta - 1) kt}$$

4° Cas limite $\beta = 0$: p tend vers 0. (AA et aa sont équidistants de $\underline{O}_0$)

Nous avons vu que l'expression de Δp est alors :

$$\Delta p = kp^2(p - 1)$$

La quadrature :

$$kt = \int_{p_0}^p \frac{dp}{p^2(p-1)}$$

donne :

$$(8) \quad kt = \frac{1}{p} - \frac{1}{p_0} - L \frac{p}{p_0} + L \frac{1-p}{1-p_0}$$

d'où :

$$pkt = 1 - p/p_0 - pLp + pLp_0 + pL(1-p) - pL(1-p_0)$$

quantité qui tend vers 1 lorsque p tend vers 0. Donc, lorsque t augmente indéfiniment, on a :

$$p \sim 1/kt \quad \text{où} \quad k = 2\mu\alpha^2(1-d) > 0$$

Asymptotiquement, la loi de décroissance de p est harmonique : p est inversement proportionnel au nombre des générations.

5° Cas limite $\beta = 1$: p tend vers 1. (AA et Aa sont équidistants de $\underline{O}$.)

L'expression de Δp est alors :

$$\Delta p = kp(p-1)^2$$

Donc, en posant $q = 1 - p$, on a :

$$\Delta q = -\Delta p = kq^2(q-1)$$

Nous sommes ramenés au cas précédent : lorsque t augmente indéfiniment, on a :

$$q = 1 - p \sim 1/kt \quad \text{où} \quad k = 2\mu\alpha^2(d+1)$$

6° Cas particulier $h = (d^2 - 1)/2d$: p tend vers 0.

Nous avons vu que l'expression de Δp est alors :

$$\Delta p = -2\mu\alpha^2hp(p-1)$$

ce qui conduit à la quadrature :

$$-2\mu\alpha^2ht = \int_{p_0}^p \frac{dp}{p(p-1)}$$

ou :

$$(9) \quad -2\mu\alpha^2ht = -L \frac{p}{p_0} + L \frac{1-p}{1-p_0}$$

On en déduit :

$$\frac{p}{p_0} = e^{\frac{2 \mu \alpha^2 h t}{1 - p_0}}$$

Donc, lorsque t augmente indéfiniment, on a :

$$p \sim \frac{p_0}{1 - p_0} e^{2 \mu \alpha^2 h t}$$

fonction exponentielle décroissante puisque h est < 0 .

*
* *

Etude particulière du cas de la dominance complète : $d = 1$.

Nous supposons tout d'abord $h \neq 0$, c'est-à-dire que les deux phénotypes ne sont pas équidistants de l'optimum $\underline{O}$. Les formules (3) deviennent alors :

(3'')

$$\begin{aligned} h &= \underline{O}/\alpha - 1 \\ \beta &= 1 \\ k &= 4 \mu \alpha^2 h \end{aligned}$$

et l'expression de la variation Δp est :

$$\Delta p = kp (p - 1)^2$$

Δp a donc le signe de k , c'est-à-dire de h .

En remplaçant l'intégration de la différence première par une quadrature, on obtient :

$$kt = \int_{p_0}^p \frac{dp}{p (p - 1)^2}$$

ou :

(10)

$$kt = \frac{1}{1 - p} - \frac{1}{1 - p_0} - L \frac{1 - p}{1 - p_0} + L \frac{p}{p_0}$$

1° Si $h > 0$, c'est-à-dire $\underline{O} > \alpha$, ce qui signifie que le phénotype A est le mieux adapté :

Δp est > 0 ; p augmente constamment et tend vers 1 lorsque t augmente indéfiniment : il y a fixation asymptotique du gène A.

La relation (10) s'écrit :

$$kt(1 - p) = 1 - \frac{1 - p}{1 - p_0} - (1 - p)L(1 - p) + (1 - p)L(1 - p_0) + (1 - p)Lp - (1 - p)Lp_0$$

Le deuxième membre tend vers 1 lorsque p tend vers 1. Donc, si t augmente indéfiniment, on a :

$$1 - p \sim 1/kt$$

La disparition du gène a s'effectue asymptotiquement suivant une loi de décroissance harmonique.

2° Si $h < 0$, c'est-à-dire $\underline{O} < \alpha$, ce qui signifie que le phénotype a est le mieux adapté :

Δp est < 0 ; p diminue constamment et tend vers 0 lorsque t augmente indéfiniment : il y a *fixation asymptotique du gène a* .

La relation (10) peut s'écrire :

$$L \frac{p}{p_0} = kt - \frac{1}{1-p} + \frac{1}{1-p_0} + L \frac{1-p}{1-p_0}$$

ou :

$$\frac{p}{p_0} = e^{kt} \cdot \frac{1-p}{1-p_0} \cdot e^{\frac{1}{1-p_0} - \frac{1}{1-p}}$$

Donc, lorsque p tend vers 0, c'est-à-dire lorsque t augmente indéfiniment, on a :

$$p \sim C e^{kt}$$

en posant :

$$C = \frac{p_0}{1-p_0} \cdot e^{\frac{p_0}{1-p_0}}$$

Comme k est < 0 , p tend alors vers 0 asymptotiquement comme la fonction exponentielle décroissante :

$$e^{kt}$$

Cas particulier : $h = 0$, c'est-à-dire : $\underline{O} = \alpha$.

Les deux phénotypes sont équidistants de l'optimum.

L'expression (2) de T donne alors :

$$T = 0$$

Il en résulte d'après (1-5) :

$$\Delta p = \mu p T = 0$$

Il n'y a plus de sélection.

La variation Δp de la fréquence du gène A d'une génération à la suivante est constamment nulle : la population est en équilibre.

Mais cet équilibre est instable : une variation quelconque de h , c'est-à-dire de l'optimum $\underline{O}$, entraînerait en effet la fixation asymptotique de l'un des gènes, comme nous venons de l'établir.

* * *

Les éléments qui interviennent dans la détermination de la position de l'optimum $\underline{O}$ sont multiples et notre intention n'est pas de les analyser ici ; mais parmi eux figure l'état du milieu environnant. Une modification dans ce milieu peut donc entraîner une variation de $\underline{O}$.

Si, initialement, la population est en équilibre, les deux gènes A et a coexistant, celui-ci est alors rompu.

Si la variation de $\underline{O}$ est faible, ou si, de façon plus précise, le nouvel optimum est tel que h reste compris dans l'intervalle :

$$(d - 1)/2 ; \quad (d + 1)/2$$

c'est-à-dire que le génotype hétérozygote Aa reste le mieux adapté, la population évoluera vers un nouvel état d'équilibre où les deux allèles coexisteront encore.

Mais, si la variation de $\underline{O}$ est telle que h devienne extérieur à cet intervalle, elle peut entraîner la disparition de l'un des allèles.

Ce phénomène peut se produire lors du déplacement d'une population nombreuse initialement en équilibre, en un lieu où les conditions de milieu sont différentes.

On peut même envisager le cas limite d'un caractère primaire à dominance complète et d'une population en équilibre instable ($d = 1$; $h = 0$). Si deux fractions importantes de cette population sont transplantées dans des milieux même très voisins, elles pourront évoluer respectivement vers la fixation de deux allèles différents, si les variations de l'optimum dues aux déplacements sont de sens contraires.

* * *

Rôle de la dominance du caractère primaire.

Proposons-nous maintenant d'analyser, à l'aide des résultats que nous venons d'établir, le rôle que joue la dominance du caractère primaire dans l'évolution de la population et en particulier son influence sur la répartition asymptotique des gènes et sur la vitesse d'évolution de la population.

Remarquons tout d'abord que, d'après l'étude précédente, les deux gènes allèles coexistent dans l'état d'équilibre si l'on a :

$$(d - 1)/2 < h < (1 + d)/2$$

c'est-à-dire si l'optimum $\underline{O}$ est compris dans l'intervalle :

$$\propto (1 + d)/2 ; \quad \propto (3 + d)/2$$

L'amplitude de cet intervalle est indépendante de d et est égale à α . Seule la valeur médiane de l'intervalle est déplacée, du fait de la dominance, d'une quantité proportionnelle à d , vers la valeur du caractère correspondant au gène dominant.

Dans le cas de la dominance complète, la remarque précédente ne semble pas justifiée ; néanmoins, on peut considérer ce cas comme un cas limite ; en effet, si $\underline{O}$ est alors compris dans l'intervalle $(\alpha ; 2\alpha)$, la population tend vers un équilibre asymptotique, mais la fréquence limite du gène A, fréquence qui, comme nous le verrons ultérieurement, est fonction du taux de dominance d , est précisément égale à 1 lorsque $d = 1$. Ceci explique que, dans le cas de la dominance complète, il y ait toujours disparition asymptotique de l'un des gènes, excepté pour la valeur α de $\underline{O}$, pour laquelle la population est en équilibre instable.

* * *

Effet de la dominance sur les valeurs asymptotiques des fréquences des gènes :

Nous supposons fixée la position $\underline{O}$ de l'optimum et nous désignerons par l la limite vers laquelle tend p lorsque t augmente indéfiniment. Etudions comment varie cette limite en fonction du taux de dominance d du caractère primaire.

La discussion précédente nous a montré que :

$$\left\{ \begin{array}{ll} \text{Si } d \leq 2h - 1 & \text{on a : } l = 1 \\ \text{Si } d \geq 2h + 1 & \text{on a : } l = 0 \\ \text{Si } 2h - 1 < d < 2h + 1 & \text{on a : } l = \beta \end{array} \right.$$

Comme $0 \leq d \leq 1$, ceci nous conduit à distinguer plusieurs cas suivant les valeurs de h , c'est-à-dire suivant la position de $\underline{O}$.

1° $h \leq -1/2$ ou $\underline{O} \leq \alpha/2$:

On a, quel que soit d : $d \geq 2h + 1$, donc il y a toujours fixation asymptotique du gène a .

2° $-1/2 < h < 0$ ou $\alpha/2 < \underline{O} < \alpha$:

$$\begin{array}{ll} \text{Si } 0 \leq d < 2h + 1 & \text{on a } l = \beta. \\ \text{Si } d \geq 2h + 1 & \text{on a } l = 1. \end{array}$$

Considérons la fonction β de d ; sa dérivée est :

(11)

$$\beta' = \frac{2h(d-h)}{(d^2 - 2dh - 1)^2}$$

Cette dérivée est < 0 .

β est par conséquent une fonction décroissante de d dans l'intervalle $(0 ; 2h + 1)$.

Nous voyons que le rôle de la dominance est de diminuer la fréquence asymptotique du gène A, et même, si le taux de dominance est assez élevé ($d \geq 2h + 1$), d'entraîner la fixation du gène a.

Le tableau suivant résume ces résultats :

d	0	$1 + 2h$	1
l	$h + \frac{1}{2}$	0	Fixation du gène a

3° Cas particulier $h = 0$ ou $\underline{O} = \alpha$: (homozygotes équidistants de $\underline{O}$).

La fréquence limite l est égale à $\beta = 1/2$ quel que soit le taux de dominance (excepté si $d = 1$).

Donc, si l'optimum $\underline{O}$ est équidistant des valeurs du caractère primaire correspondant aux deux homozygotes AA et aa, la fréquence asymptotique des gènes est indépendante du taux de dominance.

Nous avons vu d'ailleurs que dans le cas de la dominance complète la population est alors en état d'équilibre instable.

4° $0 < h \leq 1/2$ ou $\alpha \leq \underline{O} \leq 3\alpha/2$:

On a alors, quel que soit d : $2h - 1 < d < 2h + 1$.

Donc :

$$l = \beta$$

Il y a équilibre asymptotique avec coexistence des deux allèles, quel que soit d , excepté pour les deux cas limites : $d = 1$ et $\begin{cases} d = 0 \\ h = 1/2. \end{cases}$

Mais l'expression (11) montre que la fonction β de d présente dans l'intervalle $(0 ; 1)$ un minimum pour

$$d = h.$$

La dominance a donc pour effet de diminuer la fréquence asymptotique du gène A, si le taux de dominance est faible ($d < 2h$), cette diminution étant maximum pour $d = h$, valeur pour laquelle la fréquence limite est :

(12)

$$l = \frac{1}{2} + \frac{h}{1 + h^2}$$

et au contraire d'augmenter la fréquence asymptotique de A, si le taux de dominance est assez élevé ($d > 2h$), cette augmentation allant jusqu'à la fixation du gène A si la dominance est complète.

Remarquons que pour la valeur $d = 2h$, la fréquence limite du gène A est la même que s'il n'y avait pas de dominance : on a alors :

$$l = 1/2 + h$$

Le tableau suivant résume ces résultats :

d	0	h	$2h$	1
l	$\frac{1}{2} + h$	$\frac{1}{2} + \frac{h}{1+h^2}$	$\frac{1}{2} + h$	1

5° $1/2 < h < 1$ ou $3\alpha/2 < \underline{Q} < 2\alpha$:

Si le taux de dominance est faible ($d \leq 2h - 1$), il y a fixation asymptotique du gène A.

Si le taux de dominance est assez élevé ($d > 2h - 1$), il y a équilibre asymptotique entre les deux allèles, la fréquence limite du gène A étant $l = \beta$. Or, l'expression (11) montre que la fonction β de d présente encore un minimum pour $d = h$, ce minimum ayant la valeur (12). D'où le tableau suivant :

d	0	$2h - 1$	h	1
l	Fixation du gène A	1	$\frac{1}{2} + \frac{h}{1+h^2}$	1

6° $h \geq 1$ ou $\underline{Q} \geq 2\alpha$:

On a, quel que soit d : $d \leq 2h - 1$.

Il y a donc toujours fixation asymptotique du gène A.

* * *

Construisons les courbes qui représentent pour chaque valeur de d les variations de l en fonction de h . Elles se composent de deux demi-droites :

$$1) h < -(1-d)/2; \quad l = 0$$

$$2) h > (1+d)/2; \quad l = 1$$

raccordées par un arc d'hyperbole équilatère :

$$l = \frac{1}{2} - \frac{h}{d^2 - 2dh - 1}; \quad -(1-d)/2 < h < (1+d)/2$$

excepté dans les deux cas particuliers suivants :

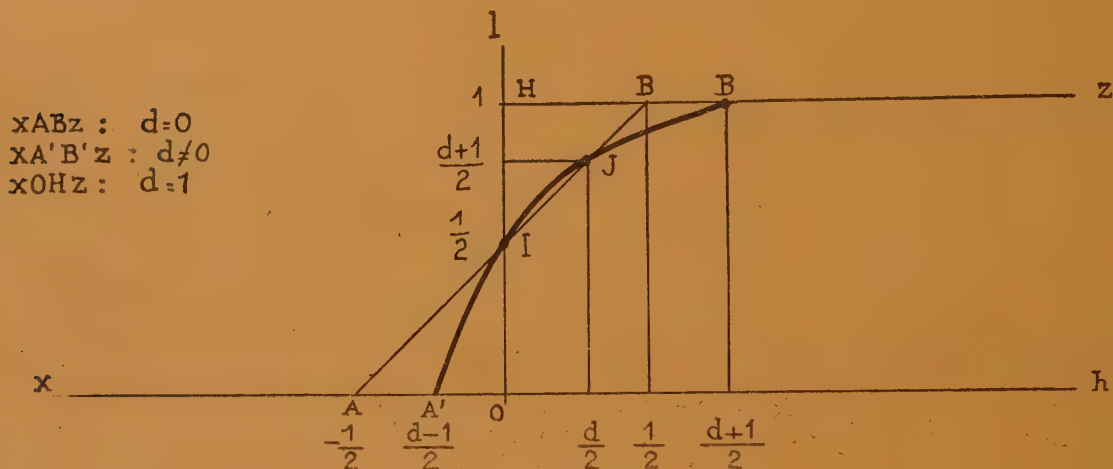
a) $d = 0$: La courbe de raccordement est le segment de droite :

$$l = 1/2 + h; \quad -(1/2) < h < 1/2$$

b) $d = 1$: La courbe se compose des deux demi-droites :

$$h < 0, \quad l = 0 \quad \text{et} \quad h > 0, \quad l = 1$$

raccordées par le segment $(0, 1)$ de l'axe des ordonnées.



Si l'on considère l'ensemble de ces courbes, on constate que tous les arcs d'hyperboles passent par le point I ($h = 0 ; l = 1/2$) et on peut déduire de leur examen les résultats que nous avons établis dans la discussion précédente.

En particulier, traçons sur le même graphique les courbes correspondant à $d = 0$, et à une valeur déterminée $d \neq 0$ du coefficient de dominance. L'arc d'hyperbole de cette dernière courbe coupe le segment de droite AB de la courbe $d = 0$ aux points :

$$I \left(h = 0 ; l = \frac{1}{2} \right) \quad \text{et} \quad J \left(h = d/2 ; l = (d + 1)/2 \right)$$

Le simple examen de ce graphique montre alors que la dominance diminue la fréquence asymptotique du gène dominant si l'optimum $\underline{O}$ est inférieur à la valeur médiane α du caractère ($h < 0$), ou s'il lui est supérieur et en diffère de plus de $\alpha d/2$ ($h > d/2$).

Au contraire, si l'optimum est supérieur à la valeur médiane mais en diffère peu ($\alpha < \underline{O} < \alpha + \alpha d/2$), la dominance augmente la fréquence asymptotique du gène dominant.

* * *

Influence de la dominance sur la rapidité d'évolution.

Proposons-nous de comparer la rapidité d'évolution d'une population où le caractère primaire présente le taux de dominance $d \neq 0$, à celle qu'aurait cette même population si le

caractère ne présentait pas de dominance. Pour que les évolutions soient comparables, il faut que la fréquence asymptotique l du gène A soit la même dans les deux cas, ce qui ne se produit que dans les quatre hypothèses suivantes que nous examinerons successivement :

- 1° $h < -1/2$: fixation asymptotique du gène a : $l = 0$;
- 2° $h > (d + 1)/2$: fixation asymptotique du gène A : $l = 1$;
- 3° $h = 0$; $l = 1/2$;
- 4° $h = d/2$; $l = (d + 1)/2$.

Toutes les autres hypothèses que l'on pourrait faire sur la valeur de h conduiraient à des valeurs différentes de l pour $d = 0$ et pour $d \neq 0$, et n'ont pas à être envisagées.

1° $h < -1/2$; fixation du gène a :

Nous avons établi page 15, que p tend asymptotiquement vers 0 comme l'exponentielle

$$\frac{\beta k t}{e}$$

Désignons par β_0 et k_0 les valeurs correspondant à l'absence de dominance. On a :

$$k_0 = 2 \mu \alpha^2 ; \quad \beta_0 = h + 1/2$$

d'où :

$$\beta_0 k_0 = 2 \mu h \alpha^2 + \mu \alpha^2$$

Or :

$$\beta k = \mu \alpha^2 (1 + 2 d h - d^2) + 2 \mu h \alpha^2$$

Donc finalement :

$$\boxed{\beta k = \beta_0 k_0 + \mu \alpha^2 d (2 h - d)}$$

expression qui montre que βk est une fonction décroissante de d ; en effet, sa dérivée :

$$(\beta k)' = 2 \mu \alpha^2 (h - d)$$

est dans le cas actuel négative quel que soit d .

La dominance augmente donc la rapidité asymptotique de fixation du gène a :

d	0	1
βk	$\mu \alpha^2 (1 + 2 h)$	$4 \mu \alpha^2 h$

2° $h > (d + 1)/2$; fixation du gène A :

Nous avons établi page 16 que $(1 - p)$ tend asymptotiquement vers 0 comme l'exponentielle :

$$\frac{(1 - \beta) k t}{e}$$

Or, on a :

$$\begin{aligned}(1 - \beta_0) k_0 &= \mu \alpha^2 - 2 h \mu \alpha^2 \\ (1 - \beta) k &= \mu \alpha^2 (1 + 2 dh - d^2) - 2 \mu \alpha^2 h\end{aligned}$$

et finalement :

$$(1 - \beta) k = (1 - \beta_0) k_0 + \mu \alpha^2 d (2 h - d)$$

expression qui montre que la fonction $(1 - \beta) k$ de d est croissante puisque sa dérivée :

$$2 \mu \alpha^2 (h - d)$$

est constamment positive dans le cas considéré.

Donc la dominance retarde la fixation du gène A .

3° $h = 0$; $l = 1/2$:

Nous avons établi page 16 que $|p - \beta|$ tend vers 0 comme l'exponentielle :

$$e^{\beta(\beta-1)kt}$$

Donc ici, $\left| p - \frac{1}{2} \right|$ tend asymptotiquement vers 0 comme l'exponentielle :

$$e^{-\frac{kt}{4}}$$

Or :

$$k = 2 \mu \alpha^2 (1 - d^2) = k_0 (1 - d^2) < k_0$$

ce qui montre que la dominance retarde asymptotiquement l'évolution de la population.

4° $h = d/2$; $l = (d + 1)/2$:

Les résultats établis page 16 montrent encore que $|p - l|$ tend vers 0 comme la fonction exponentielle :

$$e^{l(l-1)kt} = e^{(d^2-1)kt/4}$$

Or on a :

$$k = 2 \mu \alpha^2 = k_0$$

Donc, la rapidité asymptotique d'évolution de la population est la même que s'il n'y a pas de dominance.

CHAPITRE III

INFLUENCE DES MUTATIONS SUR L'EVOLUTION

Nous avons, dans les chapitres précédents, envisagé l'évolution d'une population soumise à une loi d'union sélective dans l'hypothèse où il n'y a pas de mutation entre gènes allèles.

Supposons maintenant que, par suite de mutations, une fraction u des gènes a soient, à la naissance, transformés en gènes A , et que inversement une fraction v des gènes A soient mutés en gènes a . Nous considérerons les coefficients u, v, μ , suffisamment petits pour que l'on puisse négliger $0(u^2), 0(uv), 0(\mu u)$...

Si p est la proportion des gènes A dans la génération F , p' leur fréquence dans la génération suivante F' , et si l'on désigne par p_0' la fréquence qu'auraient les gènes A dans F' s'il n'y avait pas de mutation, on a d'après ce qui précède :

$$p_0' = p(1 + \mu T); \quad 1 - p_0' = q_0' = q(1 + \mu S)$$

donc comme :

$$\begin{cases} p' = p_0'(1 - v) + (1 - p_0')u = p_0'(1 - v - u) + u \\ q' = (1 - p_0')(1 - u) + p_0'v = (1 - p_0')(1 - u - v) + v \end{cases}$$

avec l'approximation adoptée, ces quantités s'écrivent :

$$(1) \quad \begin{cases} p' = p(1 + \mu T - u - v) + u \\ q' = q(1 + \mu S - v - u) + v \end{cases}$$

et l'on a :

$$(2) \quad \Delta p = \mu p T - p v + q u$$

Or, en désignant par $P', 2Q', R'$ les fréquences des trois génotypes dans la génération F' , et par $P_0', 2Q_0', R_0'$ les fréquences qu'ils auraient si la génération F se reproduisait sans mutation, on a :

$$P_0' = p * p; \quad 2Q_0' = 2p * q; \quad R_0' = q * q$$

D'où l'on déduit :

$$\begin{cases} P' = (p * p)(1 - v)^2 + 2(p * q)(1 - v)u + (q * q)u^2 \\ 2Q' = 2(p * q)(1 - u)(1 - v) + 2(p * p)(1 - v)v + 2(q * q)(1 - u)u + 2(p * q)uv \\ R' = (q * q)(1 - u)^2 + 2(p * q)(1 - u)v + (p * p)v^2 \end{cases}$$

ou, avec l'approximation convenue :

$$\begin{cases} P' = p^2 (1 + 2 \mu T - 2 v) + 2 p q u \\ 2 Q' = 2 p q (1 + \mu T + \mu S - u - v) + 2 p^2 v + 2 q^2 u \\ R' = q^2 (1 + 2 \mu S - 2 u) + 2 p q v \end{cases}$$

Or, d'après (1) :

$$\begin{cases} p'^2 = p^2 (1 + 2 \mu T - 2 u - 2 v) + 2 p u \\ 2 p' q' = 2 p q (1 + \mu T + \mu S - 2 u - 2 v) + 2 p v + 2 q u \\ q'^2 = q^2 (1 + 2 \mu S - 2 u - 2 v) + 2 q v \end{cases}$$

La comparaison des deux groupes d'expressions établies donne immédiatement :

$$\begin{cases} P' = p'^2 \\ 2 Q' = 2 p' q' \\ R' = q'^2 \end{cases}$$

La loi de Hardy est encore vérifiée.

On peut donc encore prendre pour T l'expression établie au chapitre II (II-2 et 2') puisque l'on a :

$$T = (PI + QJ)/p = pI + qJ$$

Etudions tout d'abord le cas où le caractère primaire est sans dominance.

On obtient alors :

$$\Delta p = u - p(u + v) + \mu(1 - p)p\alpha^2(2Q/\alpha - 2p - 1)$$

ou encore en utilisant les notations (II-3') :

$$(3) \quad \Delta p = kp(p - 1)(p - \beta) + u - p(u + v)$$

On trouve pour expression de Δp un polynôme du 3^e degré, résultat analogue à celui que donne M. MALECOT dans le cas d'une population panmictique soumise à une certaine sélection (1)

Pour étudier l'évolution de notre population, nous emploierons la méthode utilisée par M. MALECOT.

L'équation (3) peut s'écrire :

$$(4) \quad \frac{\Delta p}{p(1 - p)} = k(\beta - p) + \frac{u}{p} - \frac{v}{1 - p}$$

(1) cf. G. MALECOT : « Les Mathématiques de l'Hérédité », p. 39.

Considérons la courbe (C) d'équation :

$$(C) \quad y_1 = -\frac{u}{p} + \frac{v}{1-p}$$

et la droite (D) d'équation :

$$(D) \quad y_2 = k(\beta - p)$$

et faisons varier p de 0 à 1.

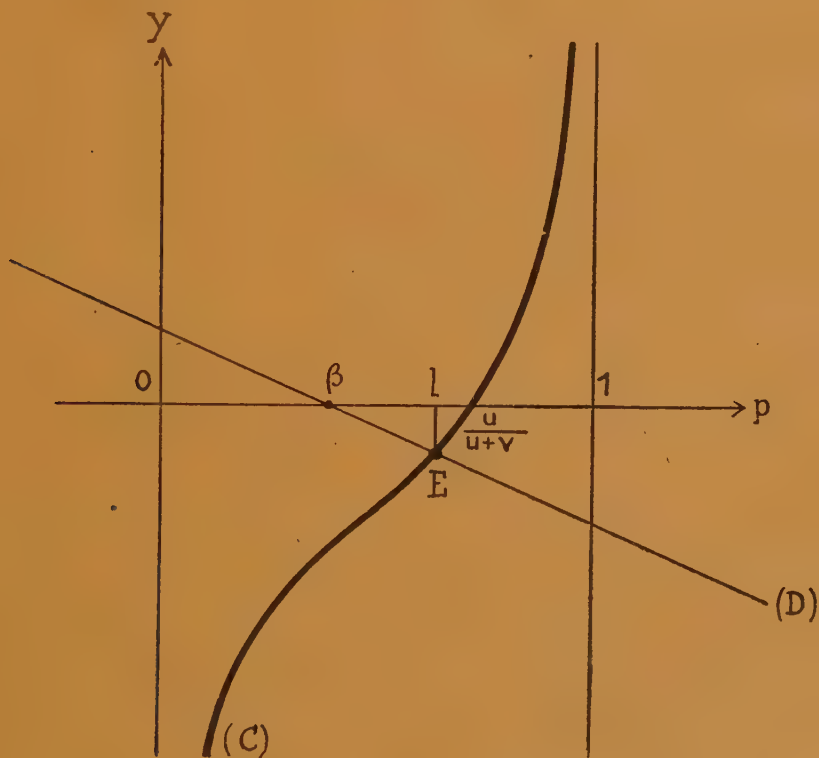
Remarquons que $y_1 = 0$ pour $p = \frac{u}{u+v}$.

D'autre part, k étant positif ($k = 2\alpha^2\mu$), le coefficient angulaire de la droite (D) est négatif : donc (D) coupe toujours la courbe (C) en un point et un seul, soit E, d'abscisse l .

On a d'après (4) :

$$\frac{\Delta p}{p(1-p)} = y_2 - y_1$$

Donc Δp a le signe de $(y_2 - y_1)$. Or :



Si $p < l$, $y_2 - y_1$ est > 0 .

Si $p > l$, $y_2 - y_1$ est < 0 .

Il en résulte que Δp a le signe de :

$$-(p - l)$$

et donc que $|p - l|$ décroît lorsqu'on passe d'une génération à la suivante.

Montrons que cette différence tend vers 0 lorsque t augmente indéfiniment, c'est-à-dire que la population tend vers un état d'équilibre stable où la fréquence des gènes A est l , et cherchons avec quelle rapidité s'établit cet état d'équilibre.

Nous affecterons de l'indice i les quantités qui se rapportent à la génération de rang i .

Le polynôme $f(p) = \Delta p$ est divisible par $(l - p)$; comme il n'a qu'une racine comprise entre 0 et 1, le polynôme du deuxième degré

$$\frac{f(p)}{l - p}$$

n'a pas de racine comprise entre 0 et 1, est toujours > 0 dans cet intervalle et présente dans l'intervalle $(p_0; l)$ une valeur minimum positive m . Posons alors :

$$z_i = p_i - l \quad \text{d'où :} \quad \Delta p_i = p_{i+1} - p_i = z_{i+1} - z_i$$

On a :

$$\frac{z_i - z_{i+1}}{z_i} > m$$

Donc :

$$\begin{array}{ll} \text{si } z_i > 0 & \text{on a } z_{i+1} < (1 - m) z_i \\ \text{si } z_i < 0 & \text{on a } z_{i+1} > (1 - m) z_i \end{array}$$

Supposons que z_{i+1} et z_i soient de même signe, hypothèse que l'on peut légitimer de la façon suivante : on a :

$$f(l) = 0$$

Donc on peut en première approximation écrire si p est voisin de l :

(5)

$$\Delta p = f'(l) (p - l)$$

et comme :

$$f'(l) = k [(l - 1)(l - \beta) + l(l - \beta) + l(l - 1)] - (u + v) = 0(\mu) + 0(u)$$

on déduit de (5) :

$$\Delta p_i = (p_i - l) \cdot 0(\mu) \quad \text{ou} \quad |\Delta p_i| < |p_i - l|$$

ce qui signifie bien que z_{i+1} et z_i sont de même signe.

On a alors dans tous les cas :

$$|z_{i+1}| < (1 - m) |z_i|$$

d'où :

$$|z_n| < |z_{n-1}| (1 - m) < |z_{n-2}| (1 - m)^2 < \dots < |z_0| (1 - m)^n$$

ce qui montre que $|z_n|$ tend vers 0 plus rapidement que l'exponentielle $(1 - m)^n$.

La population présente donc toujours un état d'équilibre asymptotique stable avec coexistence des deux gènes allèles.

Remarquons que l'égalité (5) donne une valeur approximative de m :

$$m = -f'(l)$$

et que, d'autre part, il résulte du graphique que la fréquence asymptotique du gène A satisfait aux inégalités :

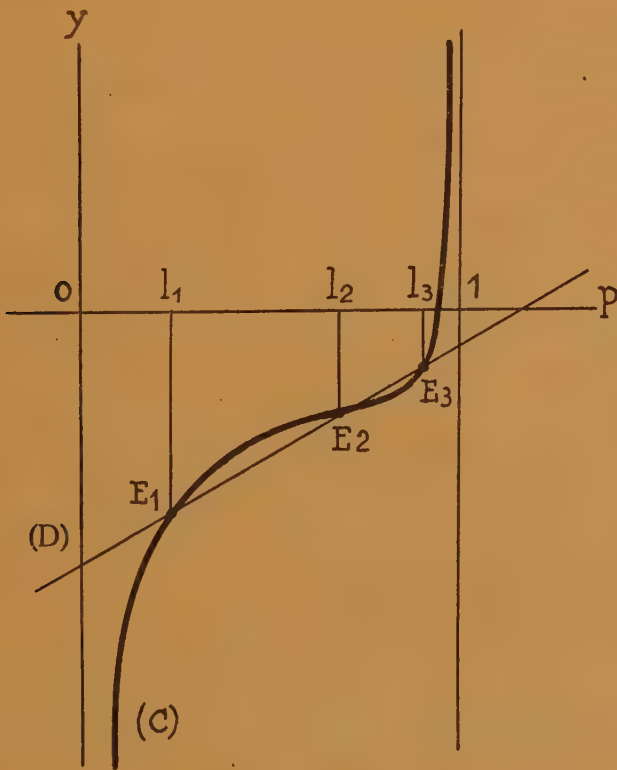
$$\beta < l < u/(u+v) \quad \text{ou} \quad u/(u+v) < l < \beta$$

ce qui signifie que cette fréquence est intermédiaire entre celle qu'aurait le gène A s'il n'y avait pas de sélection ($u/u+v$) et celle qu'il aurait s'il n'y avait pas de mutation (β , 0, ou 1, suivant la position de β par rapport à 0 et 1).

*
* *

Cas où le caractère primaire présente un taux de dominance d .

L'expression (2) s'écrit encore sous la forme (3), mais les quantités k et β ont alors les valeurs (II-3) établies page 10. En particulier, le coefficient k n'est plus toujours > 0 ; il en résulte que la droite (D) peut couper la courbe (C) en plus d'un point.



Lorsqu'il n'y a qu'un seul point d'intersection, le raisonnement que nous venons de faire dans le cas de non-dominance du caractère primaire est encore valable, et la population tend vers un état d'équilibre stable dans lequel le gène A a la fréquence l abscisse du point d'intersection de (C) et de (D).

S'il y a trois points d'intersection E_1, E_2, E_3 d'abscisses respectives :

$$l_1 < l_2 < l_3$$

Δp ayant le signe de $y_2 - y_1$, l'examen de la figure montre que, en désignant par p_0 la fréquence initiale du gène A :

Si $p_0 < l_2$, Δp a le signe de $l_1 - p_0$.

Si $p_0 > l_2$, Δp a le signe de $l_3 - p_0$.

Il en résulte que E_1 et E_3 correspondent à des états d'équilibre stables et E_2 à un état d'équilibre instable.

Si, par exemple, $p_0 < l_2$, en désignant par m le minimum de la quantité positive

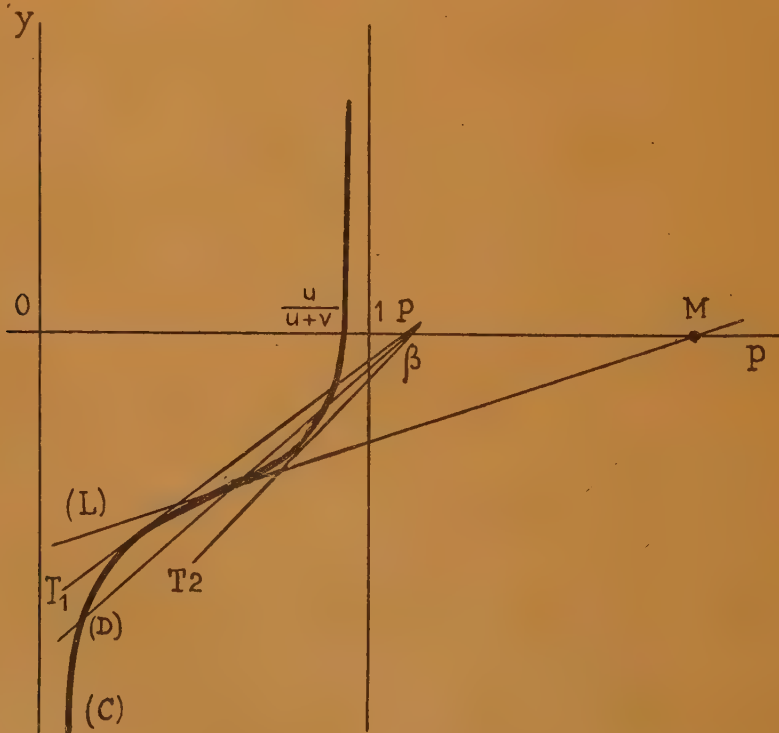
$$\frac{\Delta p}{l_1 - p}$$

dans l'intervalle (p_0, l_1) , on verrait, comme précédemment que $|l_1 - p|$ tend vers 0 lorsque le nombre n des générations augmente indéfiniment, plus rapidement que $(1 - m)^n$.

Et de même, si $p_0 > l_2$, $|l_3 - p|$ tend vers 0 plus rapidement que $(1 - m)^n$, m désignant alors le minimum de la quantité positive

$$\frac{\Delta p}{l_3 - p}$$

dans l'intervalle $(p_0 ; l_3)$.



Il est intéressant de déterminer dans quelles hypothèses la population peut présenter ces trois états d'équilibre asymptotiques, c'est-à-dire à quelles conditions doivent satisfaire les coefficients u, v, k, β , pour que la droite (D) coupe la courbe (C) en trois points.

Remarquons tout d'abord que ceci n'est possible que si le coefficient angulaire de (D) est positif, c'est-à-dire si :

$$-k > 0 \quad \text{soit :} \quad k < 0$$

Or, la discussion que nous avons faite au chapitre précédent (II, page 10), nous a montré que l'on a alors nécessairement :

$$\beta > 1$$

L'abscisse β du point P où la droite (D) coupe l'axe des abscisses est donc supérieure à 1.

D'un point P, il est possible de mener des droites qui coupent (C) en trois points, s'il existe des tangentes à la courbe (C) issues de P. Il faut pour cela que β soit inférieur à la valeur maximum de l'abscisse à l'origine des tangentes à (C), valeur maximum qui, lorsqu'elle est supérieure à 1, n'est autre que l'abscisse à l'origine de la tangente d'inflexion (L) de (C).

Cherchons donc l'abscisse du point M où la tangente d'inflexion (L) coupe l'axe des abscisses. Les expressions des dérivées première et seconde de y_1 :

$$y'_1 = \frac{u}{p^2} + \frac{v}{(1-p)^2}; \quad y''_1 = \frac{-2u}{p^3} + \frac{2v}{(1-p)^3}$$

permettent de calculer l'abscisse X_i du point d'inflexion :

$$X_i = \frac{u^{1/3}}{u^{1/3} + v^{1/3}}$$

et le coefficient angulaire de la tangente d'inflexion (L) :

$$(u^{1/3} + v^{1/3})^3$$

Posons :

$$\lambda = (u/v)^{1/3}$$

La droite (L) a alors pour équation :

$$y + (1 + \lambda) \left(\frac{u}{\lambda} - v \right) = \left(p - \frac{\lambda}{1 + \lambda} \right) \left(\frac{u}{\lambda^2} + v \right) (1 + \lambda)^2 = \left(p - \frac{\lambda}{1 + \lambda} \right) (1 + \lambda)^3 v$$

Et l'expression de l'abscisse du point M où elle coupe l'axe des abscisses est, toutes réductions faites :

$$X_M = \frac{2\lambda - 1}{\lambda + 1}$$

La condition pour que du point P on puisse mener des tangentes à la courbe (C) est donc :

$$\beta < X_M \quad \text{ou} \quad \lambda(2 - \beta) > \beta + 1$$

ce qui exige, puisque λ est > 0 : $\beta < 2$ et l'on obtient finalement :

$$(6) \quad u > v \left[\frac{1+\beta}{2-\beta} \right]^3 \quad \text{avec} \quad 1 < \beta < 2$$

La condition (6) étant réalisée, désignons par t_1 et t_2 les coefficients angulaires des tangentes menées de P à la courbe (C). La droite (D) coupe (C) en trois points si l'on a :

$$(7) \quad t_1 < -k < t_2$$

En résumé, donc, si l'ensemble des conditions (6) et (7) est réalisé, la population a trois équilibres asymptotiques possibles ; dans tous les autres cas, il y a un équilibre asymptotique parfaitement défini, qui ne dépend pas de la fréquence initiale p_0 .

Examinons ce que signifient, *qualitativement*, les conditions (6) et (7). Les inégalités (7) exigent évidemment que k soit < 0 , donc (cf. II, page 10) :

$$h < (d^2 - 1)/2d \quad \text{d'où :} \quad \underline{0} < \alpha (d + 1)/2$$

ce qui exprime que le génotype le mieux adapté est l'homozygote récessif aa .

Si l'on remarque que, lorsque β croît de 1 à 2, la quantité :

$$[(1 + \beta)/(2 - \beta)]^3$$

croît de 8 à $+\infty$, on voit que la condition (6) signifie que les coefficients de mutation u et v sont très différents l'un de l'autre, et que le gène dominant, c'est-à-dire le moins bien adapté, est très favorisé par rapport à son allèle récessif par les mutations.

Par ailleurs, si u est grand par rapport à v , l'abscisse X_i du point d'inflexion est voisine de 1. Or, on a toujours :

$$l_s > X_i$$

Donc, lorsque la population présente trois équilibres asymptotiques possibles, l'un d'eux correspond à un état voisin de la fixation du génotype AA , c'est-à-dire du génotype le plus mal adapté.

* * *

Cas particulier de la dominance complète :

On a alors : $\beta = 1$, d'où :

$$(8) \quad \Delta p = u - p(u + v) + kp(1 - p)^2$$

avec :

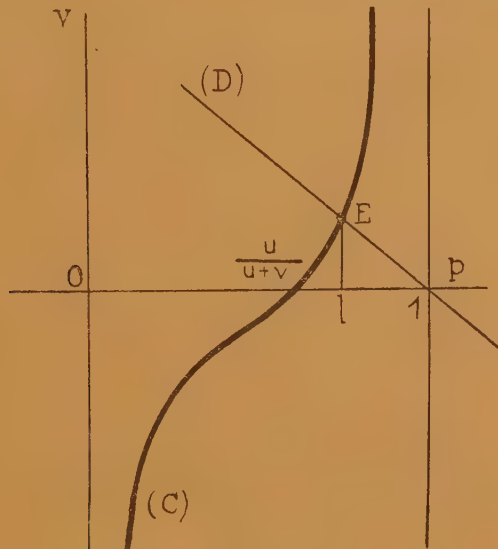
$$h = 0/\alpha - 1 ; \quad k = 4\mu\alpha^2h$$

La condition (6) s'écrit dans ce cas :

$$u > 8v$$

Ce n'est donc que si la proportion des gènes A mutés en a est plus de huit fois plus faible que celle des gènes a mutés en A que l'équilibre asymptotique (sous réserve que la condition (7) soit de plus réalisée) dépend de la fréquence initiale p_0 .

Supposons $u < 8v$; l'équilibre asymptotique est alors parfaitement défini quel que soit p_0 .



Soit $k \neq 0$. — L'examen du graphique montre que la fréquence asymptotique du gène A, abscisse l du point E d'intersection de la droite (D) et de la courbe (C) est telle que :

$$\begin{aligned} l &> u/(u+v) \quad \text{si } k > 0, \quad \text{soit : } \underline{Q} > \alpha \\ l &< u/(u+v) \quad \text{si } k < 0, \quad \text{soit : } \underline{Q} < \alpha \end{aligned}$$

Cas particulier : $k = 0$, c'est-à-dire : $\underline{Q} = \alpha$.

On a alors : $T = 0$ et l'expression (8) de Δp devient :

$$\Delta p = u - p(u+v)$$

ou encore en posant :

$$r = \frac{u}{u+v}$$

(9)

$$\Delta p = (u+v)(r-p)$$

La population est en équilibre si $p = r$, et Δp a le signe de $(r - p)$, donc :

$$\text{si } p < r, \quad \Delta p \text{ est } > 0$$

$$\text{si } p > r, \quad \Delta p \text{ est } < 0$$

Il en résulte que l'équilibre est stable. Pour voir comment la population tend vers cet état d'équilibre, assimilons l'équation aux différences finies (9) à l'équation différentielle :

$$\frac{dp}{dt} = (u + v) (r - p)$$

On obtient :

$$t = \int_{p_0}^p \frac{1}{u + v} \cdot \frac{dp}{r - p} = - \frac{1}{u + v} L \left| \frac{r - p}{r - p_0} \right|$$

D'où :

$$|r - p| = C e^{-t(u + v)}$$

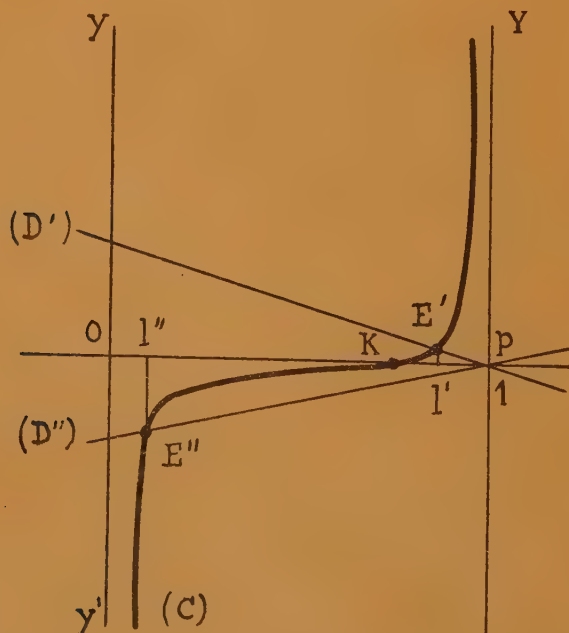
en posant :

$$C = |r - p_0|$$

La différence $|r - p|$ tend donc vers 0 comme l'exponentielle décroissante :

$$e^{-t(u + v)}$$

Remarque. — Nous avons été amenés, au chapitre II, à considérer le cas particulier $h = 0$, et nous avons montré que lorsque le caractère primaire est à dominance complète, la population est alors constamment en état d'équilibre instable, une variation très petite quelconque de l'optimum pouvant entraîner une évolution dans un sens ou dans l'autre et finalement la disparition de l'un des allèles.



L'étude présente montre au contraire que, s'il y a des mutations, la population évolue vers un état d'équilibre parfaitement défini (tout au moins si $u < 8v$), stable, même dans le cas $h = 0$.

Toutefois, supposons les coefficients u et v très petits : la courbe (C) est alors voisine du contour $y'OPY$, en désignant par P le point (1, 0) et par PY la parallèle à l'axe des ordonnées.

Pour $h = 0$, l'équilibre correspond au point K d'abscisse r où (C) coupe l'axe des abscisses

Si l'optimum $\underline{O}$ subit une légère variation, le coefficient angulaire ($-k$) de la droite (D) cesse d'être nul ; l'équilibre correspond alors à un point E qui, par suite de l'aplatissement de la courbe (C) peut être très éloigné du point K.

En particulier, deux variations faibles de l'optimum à partir de la valeur $\underline{O} = \alpha$ ($k = 0$), mais de sens contraires, conduisent à des positions d'équilibre très différentes ; pour l'une, E', la fréquence limite l' est voisine de 1 ; pour l'autre, E'', elle est voisine de 0.

CHAPITRE IV

PASSAGE D'UNE GÉNÉRATION A LA SUIVANTE

VALEURS ASYMPTOTIQUES DES PROBABILITÉS D'ASSOCIATION DES GÈNES

Nous supposerons dans ce chapitre et dans le suivant que le caractère primaire dépend de plusieurs facteurs. Soit n le nombre de ceux-ci ; les individus peuvent alors présenter 3^n génotypes différents.

Nous ferons l'hypothèse que les contributions des différents facteurs au caractère primaire sont additives, c'est-à-dire que le caractère primaire ne présente pas le phénomène d'épistacy.

Nous établirons les formules utiles dans le cas où $n = 3$ (27 génotypes), les résultats obtenus étant facilement généralisés au cas où n est quelconque.

Notations. — Nous désignerons par :

(H), (H'), (H'') les différents facteurs.

H, H', H'' les contributions des différents facteurs au caractère primaire.

$i, j, k; i', j', k'; i'', j'', k''$ les valeurs des aléatoires H, H', H'' correspondant aux trois génotypes relatifs à chaque facteur (AA, Aa, aa ; A'A', A'a', a'a' ; A''A'', A''a'', a''a'').

A, a ; A', a' ; A'', a'' les états allèles possibles des gènes correspondant aux différents facteurs.

P, 2 Q, R ; P', 2 Q', R' ; P'', 2 Q'', R'' les fréquences dans la population en équilibre, des génotypes relatifs aux trois phases possibles de chaque facteur considéré isolément.

(111), (211), (311), (121),... les fréquences des génotypes AAA'A'A''A'', AaA'A'A''A'', aaA'A'A''A'', AAA'a'A''A'', etc...

(111), (311), (131), (331),... les fréquences d'association des gènes : A, A', A'' ; a, A', A'' ; A, a', A'' ; a, a', A'' ; etc...

(111), (121),... (111),... les fréquences d'association des gènes et des génotypes : A, A'A', A''A'' ; A, A'a', A''A'' ;... AA, A', A''A''...

x : la valeur centrée du caractère primaire.

y : la valeur centrée du caractère secondaire.

V : la variance du caractère primaire.

$x_{111}, y_{111},...$ les moyennes de x et y pour les individus présentant le génotype AAA'A'A''A'', etc...

$x_{111}, y_{111}, x_{311}, y_{311}, \dots$ les moyennes de x et de y pour une catégorie d'individus dont la fréquence représente la fréquence d'association des différents gènes. Par exemple, x_{111} est la moyenne de x pour une catégorie comprenant les homozygotes $AAA'A'A''A''$, la moitié des individus de génotypes : $AAA'a'A''A''$; $AAA'A'a''A''$ et $AaA'A'A''A''$; le quart des individus de génotypes $AAA'a'A''a''$, $AaA'A'A''a''$ et $AaA'a'A''A''$; et le huitième des individus de génotypes $AaA'a'A''a''$, etc...

I, J, K ; I', J', K' ; I'', J'', K'' : les valeurs moyennes de x dans une population en équilibre lorsqu'on classe les individus en trois catégories suivant les phases du premier facteur, du deuxième ou du troisième, indépendamment les uns des autres.

Le symbole 0 dans les notations précédentes signifie que le facteur correspondant peut avoir une phase quelconque. Par exemple (130) est la probabilité d'association des gènes A et a' ; (201) est la probabilité du génotype $AaA''A''$.

Nous poserons de plus dans la population en équilibre :

$$\begin{aligned} p &= (100) ; & q &= (300) ; & p' &= (010) ; & q' &= (030) ; & p'' &= (001) ; & q'' &= (003) \\ \mathcal{C} &= x_{100} ; & \mathcal{S} &= x_{300} ; & \mathcal{C}' &= x_{010} ; & \mathcal{S}' &= x_{030} ; & \mathcal{C}'' &= x_{001} ; & \mathcal{S}'' &= x_{003} \\ T &= y_{100} ; & S &= y_{300} ; & T' &= y_{010} ; & S' &= y_{030} ; & T'' &= y_{001} ; & S'' &= y_{003} \end{aligned}$$

Enfin, au cours du premier paragraphe du présent chapitre, nous désignerons par F et F' deux générations successives, les probabilités d'association des gènes et des génotypes seront représentées dans F par les symboles non accentués (111), (111),... et dans F' par les symboles accentués (111)', (111)'...

* * *

Passage d'une génération à la suivante :

Classons dans une génération (F) les individus suivant leur génotype en 27 catégories. La fréquence du génotype $AAA'A'A''A''$ par exemple dans la génération suivante F' est, d'après les propriétés de distributivité et de commutativité du produit symbolique, donnée par le développement du carré symbolique de la quantité :

$$(111) + (112)/2 + (121)/2 + (211)/2 + (122)/4 + (212)/4 + (221)/4 + (222)/8$$

c'est-à-dire le carré symbolique de (111) fréquence d'association des gènes A, A', A'' dans F. Donc, en désignant par y_{111} la moyenne définie par :

$$\begin{aligned} (111) y_{111} &= y_{111} (111) + y_{112} (112)/2 + y_{121} (121)/2 + y_{211} (211)/2 \\ &\quad + y_{122} (122)/4 + y_{212} (212)/4 + y_{221} (221)/4 + y_{222} (222)/8 \end{aligned}$$

on a :

$$(111)' = (111) * (111) = (111)^2 (1 + 2 \mu y_{111})$$

et on obtiendrait de même les expressions analogues :

$$\begin{aligned}(112)' &= 2 (111) * (113) \\ (122)' &= 2 (111) * (133) + 2 (131) * (113) \\ (222)' &= 2 (111) * (333) + 2 (113) * (331) + 2 (131) * (313) + 2 (133) * (311) \\ &\text{etc...}\end{aligned}$$

Cherchons alors les probabilités d'association des gènes dans la génération F'. On a :

$$(111)' = (111)' + \frac{(112)' + (121)' + (211)'}{2} + \frac{(122)' + (212)' + (221)'}{4} + \frac{(222)'}{8}$$

d'où :

$$\begin{aligned}(111)' &= (111) * (111) + (111) * (113) + (111) * (131) + (111) * (311) \\ &\quad + (111) * (133)/2 + (131) * (113)/2 + (111) * (313)/2 + (311) * (113)/2 \\ &\quad + (111) * (331)/2 + (131) * (311)/2 + (111) * (333)/4 + (113) * (331)/4 \\ &\quad + (131) * (313)/4 + (133) * (311)/4\end{aligned}$$

En tenant compte des relations :

$$\begin{aligned}(110) &= (111) + (113); & (130) &= (131) + (133); & (310) &= (311) + (313) \\ (330) &= (331) + (333); & (100) &= (110) + (130); & (300) &= (310) + (330) \\ & & \text{et } (100) + (300) &= (000) = 1\end{aligned}$$

on obtient finalement :

$$(1) \quad (111)' = \frac{(110) * (001) + (101) * (010) + (011) * (100) + (111) * (000)}{4}$$

Remarquons que le dernier produit $(111) * (000)$ qui représente la probabilité pour que dans un couple pris au hasard l'un des conjoints provienne de la catégorie d'individus de fréquence (111) , n'est pas égal à (111) , par suite de la loi d'union adoptée : c'est par là que les résultats obtenus diffèrent de ceux que M. MALÉCOT a établis dans le cas du random mating et dans un cas d'homogamie. ⁽¹⁾

On établirait de la même manière l'expression :

$$(1') \quad (113)' = \frac{(110) * (003) + (103) * (010) + (013) * (100) + (113) * (000)}{4}$$

(1) Cf. G. MALÉCOT : « *Théorie mathématique de l'hérédité mendélienne généralisée* ». Thèse, Paris 1939, p. 27.

Nous avons alors :

$$\begin{aligned}(110)' &= (111)' + (113)' \\ &= \frac{(110) * (000) + (100) * (010) + (010) * (100) + (110) * (000)}{4}\end{aligned}$$

et finalement :

$$(2) \quad (110)' = \frac{(110) * (000) + (100) * (010)}{2}$$

On établirait de même l'expression analogue :

$$(2') \quad (130)' = \frac{(130) * (000) + (100) * (030)}{2}$$

Enfin :

$$\begin{aligned}(100)' &= (130)' + (110)' \\ &= \frac{(130) * (000) + (110) * (000) + (100) * (030) + (100) * (010)}{2}\end{aligned}$$

et :

$$(3) \quad (100)' = (100) * (000)$$

La proportion de chaque gène varie donc au cours des générations, contrairement à ce qui se produit dans le cas du random mating ou dans le cas d'homogamie étudié par M. MALÉCOT.

Les équations (1), (2), (3) donnent la variation d'une génération à la suivante des fréquences des gènes et de leurs fréquences d'association, à condition de pouvoir expliciter les produits symboliques, c'est-à-dire de connaître pour chaque génération les quantités y_{100} , y_{110} ,... qui représentent les valeurs moyennes du caractère secondaire lorsque les gènes occupant un ou plusieurs loci sont déterminés.

*
* * *

Population en équilibre. - Valeurs asymptotiques des probabilités d'association des gènes.

Nous considérerons comme un fait expérimental que la population tend vers un état d'équilibre, c'est-à-dire que les valeurs moyennes y_{100} , y_{110} ,... ont des limites stables.

Supposons donc la population en équilibre et étudions alors les probabilités des différents gènes ainsi que leurs probabilités d'association.

L'équation (3) donne :

$$(100) = (100) * (000) = (100) (1 + \mu y_{100})$$

ou encore avec les notations adoptées :

$$p = p (1 + \mu T)$$

ce qui exige, avec l'approximation convenue :

$$(4) \quad T = 0 (\mu)$$

De même, l'équation (2) qui s'écrit :

$$(110)' = \frac{(110) (1 + \mu y_{110}) + (100) (010) (1 + \mu y_{010} + \mu y_{100})}{2}$$

donne, à l'équilibre, en tenant compte de (4) :

$$(110) = \frac{(110) (1 + \mu y_{110}) + pp'}{2}$$

et finalement :

$$(5) \quad (110) = pp' (1 + \mu y_{110}) + 0 (\mu^2)$$

Enfin, l'équation (1) qui s'écrit :

$$4 (111)' = (110) (001) (1 + \mu y_{110} + \mu y_{001}) + (101) (010) (1 + \mu y_{101} + \mu y_{010}) \\ + (011) (100) (1 + \mu y_{011} + \mu y_{100}) + (111) (1 + \mu y_{111})$$

donne, à l'équilibre, en tenant compte de (4) et (5) :

$$4 (111) = pp'p'' [(1 + \mu y_{110})^2 + (1 + \mu y_{101})^2 + (1 + \mu y_{011})^2 + (111) (1 + \mu y_{111})]$$

et l'on trouve finalement en négligeant $0 (\mu^2)$:

$$(6) \quad (111) = pp'p'' [1 + \frac{\mu}{3} (2 y_{110} + 2 y_{101} + 2 y_{011} + y_{111})]$$

Les relations établies pages 40 et 41 nous permettent alors de calculer les fréquences limites d'association des phases des différents facteurs. Par exemple :

$$(111) = (111)^2 (1 + 2 \mu y_{111})$$

d'où :

$$(7) \quad (111) = p^2 p'^2 p''^2 [1 + \frac{4 \mu}{3} (y_{110} + y_{101} + y_{011} + 2 y_{111})]$$

De même, pour deux facteurs, on a, par exemple :

$$(110) = (110)^2 (1 + 2 \mu y_{110})$$

ou :

(8)

$$(110) = p^2 p'^2 (1 + 4 \mu y_{110})$$

Enfin, si l'on considère les phases d'un seul facteur, on a, par exemple :

$$(100) = (100)^2 (1 + 2 \mu y_{100})$$

ou :

(9)

$$(100) = p^2$$

Nous retrouvons évidemment, dans le cas de la population en équilibre, la loi de Hardy qui, comme nous l'avons démontré page 6, est vérifiée à chaque génération.

Les expressions (7), (8) et (9) nous montrent que, bien que la loi de Hardy soit vérifiée comme dans le cas du random mating, les probabilités d'association entre les phases des différents facteurs ne sont pas celles que l'on obtient dans ce cas : *les facteurs ne sont pas asymptotiquement indépendants.*

Nous avons établi des formules donnant les fréquences d'association des gènes ou des facteurs, mais celles-ci contiennent les quantités $y_{111}, y_{110}, \dots$ Pour trouver une valeur approchée de ces fréquences, nous allons maintenant déterminer les parties principales de ces valeurs moyennes du caractère secondaire.

Remarque. — La condition que nous avons imposée à x et à y de désigner les valeurs centrées des caractères primaire et secondaire nous conduirait, si la population n'était pas en équilibre, à modifier d'une génération à l'autre l'origine des échelles de mesure de ces caractères, et en particulier la valeur de l'optimum $\underline{Q}$ varierait, de ce fait, à chaque génération. C'est pour cette raison que nous n'avons pas adopté au cours des chapitres précédents où nous avons étudié une population en évolution, les notations utilisées dans le présent chapitre.

Les valeurs $i, j, k; i', j', k'; \dots$ des contributions des différents facteurs (H), (H'),... au caractère primaire, peuvent évidemment être mesurées à partir d'origines arbitraires pourvu que la moyenne $\underline{M}(x)$ soit nulle. Nous choisirons donc pour mesurer H, H',... des origines telles que l'on ait :

$$\underline{M}(H) = P_i + 2 Q_j + R_k = 0;$$

etc...

$$\underline{M}(H') = P'_i + 2 Q'_j + R'_k = 0$$

ou encore avec l'approximation adoptée :

(10)

$$p^2 i + 2 p q j + q^2 k = 0$$

$$p'^2 i' + 2 p' q' j' + q'^2 k' = 0$$

etc...

On a bien alors :

$$\underline{M}(x) = \Sigma \underline{M}(H) = 0$$

puisque par hypothèse :

$$x = \Sigma H$$

Exprimons pour un individu la valeur y du caractère secondaire en fonction de la valeur x du caractère primaire ; on a par définition :

$$\begin{aligned} y &= -(x - \underline{O})^2 - M [-(x - \underline{O})^2] \\ &= -x^2 - \underline{O}^2 + 2 \underline{O}x + M(x^2) - 2 \underline{O} M(x) + \underline{O}^2 \end{aligned}$$

Et finalement :

$$(11) \quad y = -x^2 + 2 \underline{O}x + V$$

Cette formule montre que les contributions de chaque facteur au caractère secondaire ne sont pas additives : le phénomène d'épistacy apparaît en ce qui concerne le caractère secondaire bien que nous ayons supposé son absence pour le caractère primaire.

Posons alors :

$$M_{100}(x^2) = V_{100} ; \quad M_{110}(x^2) = V_{110} ; \quad \text{etc.}$$

les expressions des moyennes de y sont :

$$(12) \quad \begin{aligned} y_{100} &= 2 \underline{O}x_{100} + V - V_{100} \\ y_{110} &= 2 \underline{O}x_{110} + V - V_{110} \\ y_{111} &= 2 \underline{O}x_{111} + V - V_{111} \\ &\text{etc...} \end{aligned}$$

et nous sommes amenés à chercher les parties principales des quantités :

$$x_{100}, x_{110} \dots ; V ; V_{100}, V_{110} \dots$$

* *

Rangeons les individus de la population suivant les trois phases du facteur (H) ; les moyennes de x dans chaque catégorie sont égales à la somme des moyennes dans ces catégories des contributions des différents facteurs. On a donc :

$$\begin{aligned} x_{100} = I = i + \frac{(110) i' + (120) j' + (130) k'}{(100)} + \frac{(101) i'' + (102) j'' + (103) k''}{(100)} \\ \text{etc...} \end{aligned}$$

ce qui s'écrit en tenant compte des relations (8) et (10) :

$$(13) \quad I = i + 0(\mu) ; \quad J = j + 0(\mu) ; \quad K = k + 0(\mu)$$

D'autre part, en posant :

$$(110) = (110) + (210)/2 \quad \text{etc...}$$

on peut écrire :

(100) $x_{100} = (100)i + (200)j/2 + (110)i' + (120)j' + (130)k' + (101)i'' + (102)j'' + (103)k''$
et les relations (10) montrent encore que la somme de tous les termes, sauf les deux premiers, est de l'ordre de μ ; on a donc :

$$(100) x_{100} = Pi + Qj + 0(\mu)$$

et finalement :

$$(14) \quad \begin{aligned} x_{100} &= \mathcal{E} = (Pi + Qj)/p + 0(\mu) = pi + qj + 0(\mu) \\ x_{300} &= \mathcal{S} = (Qj + Rk)/q + 0(\mu) = pj + qk + 0(\mu) \end{aligned}$$

On écrit de la même manière :

$$(110) x_{110} = (110)i + (210)j/2 + (110)i' + (120)j'/2 + (111)i'' + (112)j'' + (113)k''$$

et toujours d'après (10), (7) et (8) :

$$(110) x_{110} = p'(Pi + Qj) + p(P'i' + Q'j') + 0(\mu)$$

ou d'après (5) :

$$x_{110} = pi + qj + p'i' + q'j' + 0(\mu)$$

et finalement :

$$(15) \quad x_{110} = \mathcal{E} + \mathcal{E}' + 0(\mu)$$

On obtiendrait de même les expressions analogues.

* * *

Pour calculer les parties principales de V , V_{100} , etc..., remarquons que la variance V s'écrit :

$$V = M(x^2) = M[x \cdot (H + H' + H'')] = \Sigma M(Hx)$$

donc :

$$(16) \quad V = \Sigma PiI + 2 QjJ + RkK = \Sigma v$$

en posant :

$$(16') \quad \begin{aligned} v &= PiI + 2 QjJ + RkK \\ v' &= P'i'I' + 2 Q'j'J' + R'k'K' \\ &\text{etc...} \end{aligned}$$

soit, d'après (13) :

$$(16'') \quad \begin{aligned} v &= Pi^2 + 2 Qj^2 + Rk^2 + 0(\mu) = PI^2 + 2 QJ^2 + RK^2 + 0(\mu) \\ &\text{etc...} \end{aligned}$$

On a de même :

$$V_{100} = \Sigma M_{100} (Hx) = M_{100} (Hx) + \Sigma' M_{100} (H'x)$$

la somme Σ' étant étendue à tous les facteurs autres que (H).

D'où :

$$\begin{aligned} pV_{100} = & PiI + QjJ + (110) i' x_{110} + (120) j' x_{120} + (130) k' x_{130} \\ & + (101) i'' x_{101} + (102) j'' x_{102} + (103) k'' x_{103} \end{aligned}$$

Or :

$$\begin{aligned} (110) x_{110} = & (110) x_{110} + (210) x_{210}/2 \\ & \text{etc...} \end{aligned}$$

et :

$$\begin{aligned} x_{110} = & i + i' + \frac{(111) i'' + (112) j'' + (113) k''}{(110)} \\ & \text{etc...} \end{aligned}$$

donc, en tenant compte de (10), (7) et (8) :

$$(17) \quad \begin{aligned} x_{110} = & i + i' + 0 (\mu) \\ x_{210} = & j + i' + 0 (\mu) \\ & \text{etc...} \end{aligned}$$

et par conséquent :

$$(18) \quad \begin{aligned} (110) x_{110} = & PP' (i + i') + QP' (j + i') + 0 (\mu) \\ (120) x_{120} = & 2 PQ' (i + j') + 2 QQ' (j + j') + 0 (\mu) \\ & \text{etc...} \end{aligned}$$

On peut donc écrire :

$$\begin{aligned} pV_{100} = & PiI + QjJ + (Pi + Qj) \Sigma' (P'i' + 2 Q'j' + R'k') \\ & + (P + Q) \Sigma' (P'i'^2 + 2 Q'j'^2 + R'k'^2) + 0 (\mu) \end{aligned}$$

ou finalement d'après (10) et (9) :

$$(19) \quad \begin{aligned} V_{100} = & pi^2 + qj^2 + \Sigma' v + 0 (\mu) \\ & \text{etc...} \end{aligned}$$

On a alors :

$$V - V_{100} = v - pi^2 - qj^2 + 0 (\mu)$$

et l'expression (12) de y_{100} devient :

$$y_{100} = 2 \underline{O} \mathcal{E} + v - (pi^2 + qj^2) + 0 (\mu)$$

ce qui conduit, puisqu'à l'équilibre $y_{100} = T = 0 (\mu)$, à la relation :

$$(20) \quad \boxed{pi^2 + qj^2 = 2 \underline{O} \mathcal{E} + v + 0 (\mu)}$$

* * *

Calculons de la même façon $V_{110}, \dots$ afin de pouvoir exprimer les quantités $y_{110}, \dots$:

$$V_{110} = \Sigma M_{110} (Hx) = M_{110} (Hx) + M_{110} (H'x) + \Sigma'' M_{110} (H''x)$$

la somme Σ'' étant étendue à tous les facteurs sauf (H) et (H'), (elle se réduit donc à $M_{110} (H''x)$ s'il n'y a que trois facteurs).

D'où :

$$(21) \quad \boxed{(110) V_{110} = (110) ix_{110} + (210) jx_{210}/2 + (110) i'x_{110} + (120) j'x_{120}/2 + \Sigma'' [(111) i''x_{111} + (112) j''x_{112} + (113) k''x_{113}]}$$

Or :

$$\begin{aligned} (111) x_{111} &= (111) x_{111} + (211) x_{211}/2 \\ &= (111) x_{111} + (121) x_{121}/2 + (211) x_{211}/2 + (221) x_{221}/4 \end{aligned}$$

Chacun des termes de la somme Σ'' figurant dans l'expression (21) est donc de la forme :

$$\begin{aligned} &i'' [(111) x_{111} + (121) x_{121}/2 + (211) x_{211}/2 + (221) x_{221}/4] \\ &+ j'' [(112) x_{112} + (122) x_{122}/2 + (212) x_{212}/2 + (222) x_{222}/4] \\ &+ k'' [(113) x_{113} + (123) x_{123}/2 + (213) x_{213}/2 + (223) x_{223}/4] \end{aligned}$$

Or :

$$x_{111} = i + i' + i'' + \frac{(1111) i''' + (1112) j''' + (1113) k'''}{(1110)} + \dots$$

etc...

(expressions qui se réduisent à $x_{111} = i + i' + i''$, etc..., s'il n'y a que trois facteurs).

Et, d'après (10), (7) et les égalités analogues, on a :

$$(22) \quad \boxed{x_{111} = i + i' + i'' + 0 (\mu)}$$

etc...

On peut alors écrire chacun des termes de la somme Σ'' de (21) de la façon suivante :

$$\begin{aligned} & P'' i'' [PP' (i + i' + i'') + PQ' (i + j' + i'') + QP' (j + i' + i'') + QQ' (j + j' + i'')] \\ & + 2 Q'' j'' [PP' (i + i' + j'') + PQ' (i + j' + j'') + QP' (j + i' + j'') + QQ' (j + j' + j'')] \\ & + R'' k'' [PP' (i + i' + k'') + PQ' (i + j' + k'') + QP' (j + i' + k'') + QQ' (j + j' + k'')] + 0 (\mu) \end{aligned}$$

ce qui d'après (10) se réduit à :

$$(P'' i''^2 + 2 Q'' j''^2 + R'' k''^2) (P + Q) (P' + Q') + 0 (\mu) = pp' v'' + 0 (\mu)$$

et dans (21) la somme Σ'' devient :

$$pp' \Sigma'' v'' + 0 (\mu) = pp' (V - v - v') + 0 (\mu)$$

Les égalités (18) donnent d'autre part :

$$(110) ix_{110} = [PP' (i + i') + PQ' (i + j')] i + 0 (\mu)$$

$$(210) jx_{210}/2 = [QP' (j + i') + QQ' (j + j')] j + 0 (\mu)$$

$$(110) i'x_{110} = [PP' (i + i') + QP' (j + i')] i' + 0 (\mu)$$

$$(120) j'x_{120}/2 = [PQ' (i + j') + QQ' (j + j')] j' + 0 (\mu)$$

L'équation (21) devient donc :

$$\begin{aligned} (110) V_{110} &= (P + Q) (P' i'^2 + Q' j'^2) + (P' + Q') (Pi^2 + Qj^2) + 2 (Pi + Qj) (P' i' + Q' j') \\ &+ pp' (V - v - v') + 0 (\mu) \end{aligned}$$

d'où en tenant compte de (5) et (14) :

$$(21') \quad V_{110} = pi^2 + qj^2 + p'i'^2 + q'j'^2 + 2 \mathcal{E}\mathcal{E}' + V - v - v' + 0 (\mu)$$

*
* *

L'expression (12) de y_{110} s'écrit alors d'après (15) et (21') :

$$y_{110} = 2 \underline{Q} (\mathcal{E} + \mathcal{E}') + v + v' - 2 \mathcal{E}\mathcal{E}' - (pi^2 + qj^2) - (p'i'^2 + q'j'^2) + 0 (\mu)$$

et la relation (20), réalisée lorsque la population est en équilibre, nous conduit à l'expression simple :

$$(23) \quad y_{110} = -2 \mathcal{E}\mathcal{E}' + 0 (\mu)$$

et les formules (5) et (8) s'écrivent :

$$\begin{aligned} (24) \quad (110) &= pp' (1 - 2 \mu \mathcal{E}\mathcal{E}') + 0 (\mu^2) \\ (110) &= p^2 p'^2 (1 - 8 \mu \mathcal{E}\mathcal{E}') + 0 (\mu^2) \end{aligned}$$

On établirait de même les probabilités d'association des autres gènes deux à deux :

$$(24) \quad \begin{array}{l} (130) = pq' (1 - 2\mu \mathcal{E} \mathcal{S}') + 0(\mu^2) \\ (310) = qp' (1 - 2\mu \mathcal{S} \mathcal{E}') + 0(\mu^2) \\ (330) = qq' (1 - 2\mu \mathcal{S} \mathcal{S}') + 0(\mu^2) \\ (101) = pp'' (1 - 2\mu \mathcal{E} \mathcal{E}'') + 0(\mu^2) \\ (103) = pq'' (1 - 2\mu \mathcal{E} \mathcal{S}'') + 0(\mu^2) \\ \text{etc...} \end{array}$$

ainsi que les probabilités d'association des phases homozygotes des facteurs deux à deux :

$$(24'') \quad \begin{array}{l} (130) = p^2 q'^2 (1 - 8\mu \mathcal{E} \mathcal{S}') + 0(\mu^2) \\ (310) = q^2 p'^2 (1 - 8\mu \mathcal{S} \mathcal{E}') + 0(\mu^2) \\ (330) = q^2 q'^2 (1 - 8\mu \mathcal{S} \mathcal{S}') + 0(\mu^2) \\ (101) = p^2 p''^2 (1 - 8\mu \mathcal{E} \mathcal{E}'') + 0(\mu^2) \\ (103) = p^2 q''^2 (1 - 8\mu \mathcal{E} \mathcal{S}'') + 0(\mu^2) \\ \text{etc...} \end{array}$$

Pour calculer les probabilités d'association deux à deux des phases des facteurs, lorsqu'une au moins de celles-ci est hétérozygote, remarquons que l'on a :

$$(120) = (100) - (110) - (130) \quad \text{et} \quad (220) = (200) - (210) - (230)$$

donc :

$$\begin{aligned} (120) &= p^2 - p^2 p'^2 (1 - 8\mu \mathcal{E} \mathcal{E}') - p^2 q'^2 (1 - 8\mu \mathcal{E} \mathcal{S}') + 0(\mu^2) \\ &= p^2 [1 - p'^2 - q'^2 + 8\mu \mathcal{E} (p'^2 \mathcal{E}' + q'^2 \mathcal{S}')] + 0(\mu^2) \end{aligned}$$

Or, si l'on tient compte de :

$$1 - p'^2 - q'^2 = 2 p' q' \quad \text{et} \quad p' \mathcal{E}' + q' \mathcal{S}' = 0$$

on a :

$$p'^2 \mathcal{E}' + q'^2 \mathcal{S}' = p' (1 - q') \mathcal{E}' + q' (1 - p') \mathcal{S}' = -p' q' (\mathcal{E}' + \mathcal{S}')$$

et on obtient finalement :

$$(25) \quad (120) = 2 p^2 p' q' [1 - 4\mu \mathcal{E} (\mathcal{E}' + \mathcal{S}')] + 0(\mu^2)$$

ainsi que les expressions analogues :

$$\begin{aligned}
 (25') \quad & (210) = 2 p q p'^2 [1 - 4 \mu (\mathcal{E} + \mathcal{I}) \mathcal{E}'] + 0 (\mu^2) \\
 & (320) = 2 q^2 p' q' [1 - 4 \mu \mathcal{I} (\mathcal{E}' + \mathcal{I}')] + 0 (\mu^2) \\
 & (230) = 2 p q q'^2 [1 - 4 \mu (\mathcal{E} + \mathcal{I}) \mathcal{I}'] + 0 (\mu^2) \\
 & (201) = 2 p q p''^2 [1 - 4 \mu (\mathcal{E} + \mathcal{I}) \mathcal{E}''] + 0 (\mu^2) \\
 & \text{etc...}
 \end{aligned}$$

On a de même :

$$\begin{aligned}
 (220) &= 2 p q - 2 p q p'^2 [1 - 4 \mu \mathcal{E}' (\mathcal{E} + \mathcal{I})] - 2 p q q'^2 [1 - 4 \mu \mathcal{I}' (\mathcal{E} + \mathcal{I})] \\
 &= 2 p q [2 p' q' - 4 \mu (\mathcal{E} + \mathcal{I}) (\mathcal{E}' + \mathcal{I}') p' q'] + 0 (\mu^2)
 \end{aligned}$$

et finalement :

$$(26) \quad (220) = 4 p q p' q' [1 - 2 \mu (\mathcal{E} + \mathcal{I}) (\mathcal{E}' + \mathcal{I}')] + 0 (\mu^2)$$

ainsi que les équations analogues :

$$\begin{aligned}
 (26') \quad & (202) = 4 p q p'' q'' [1 - 2 \mu (\mathcal{E} + \mathcal{I}) (\mathcal{E}'' + \mathcal{I}'')] + 0 (\mu^2) \\
 & (022) = 4 p' q' p'' q'' [1 - 2 \mu (\mathcal{E}' + \mathcal{I}') (\mathcal{E}'' + \mathcal{I}'')] + 0 (\mu^2)
 \end{aligned}$$

*
* *

Si nous caractérisons les probabilités d'association entre gènes ou entre phases des différents facteurs par les coefficients f_{ij} de Fisher, c'est-à-dire si nous posons :

$$\begin{aligned}
 (11) &= PP' (1 + f_{11}); & (12) &= 2 PQ' (1 + f_{12}); & (13) &= PR' (1 + f_{13}) \\
 &\text{etc...}
 \end{aligned}$$

les expressions (24), (25) et (26) donnent immédiatement les valeurs de ces coefficients :

$$\begin{aligned}
 (27) \quad & f_{11} = -8 \mu \mathcal{E} \mathcal{E}' \\
 & f_{12} = -4 \mu \mathcal{E} (\mathcal{E}' + \mathcal{I}') \\
 & f_{22} = -2 \mu (\mathcal{E} + \mathcal{I}) (\mathcal{E}' + \mathcal{I}') \\
 & \text{etc...}
 \end{aligned}$$

Il est remarquable que l'expression des coefficients de Fisher concernant les phases homozygotes de deux facteurs soit la même que celle qu'a adoptée M. Fisher et qu'a établie M. MALÉCOT dans le cas d'homogamie qu'ils ont envisagé (1). Par contre, les expressions des coefficients concernant les phases hétérozygotes des facteurs sont totalement différentes.

(1) cf. G. MALÉCOT : Thèse, Paris 1939, p. 6 et 78.

* * *

La méthode que nous avons utilisée pour établir les probabilités d'association des gènes et des phases deux à deux, permet de calculer les probabilités d'association trois à trois.

On a en effet d'après (12) :

$$y_{111} = 2 \underline{Q}x_{111} + V - V_{111}$$

Or :

$$(111) x_{111} = \Sigma [(111) i + (211) j/2]$$

la somme Σ étant étendue aux trois facteurs (H), (H') et (H'').

D'où :

$$(111) x_{111} = \Sigma [(111) + \frac{(112)}{2} + \frac{(121)}{2} + \frac{(122)}{4}] i + \Sigma [\frac{(211)}{2} + \frac{(221)}{4} + \frac{(212)}{4} + \frac{(222)}{8}] j$$

et d'après (7) et (10) :

$$\begin{aligned} (111) x_{111} &= \Sigma (p^2 i + pqj) (p'^2 p''^2 + p'^2 p'' q'' + p' q' p''^2 + p' q' p'' q'') + 0 (\mu) \\ &= \Sigma p' p'' (Pi + Qj) = pp' p'' (\mathcal{E} + \mathcal{E}' + \mathcal{E}'') + 0 (\mu) \end{aligned}$$

et :

(28)

$$\boxed{x_{111} = \mathcal{E} + \mathcal{E}' + \mathcal{E}'' + 0 (\mu) \text{ etc...}}$$

— De même en écrivant :

$$V_{111} = \Sigma M_{111} (Hx)$$

et en nous bornant au cas de trois facteurs :

$$V_{111} = M_{111} (Hx) + M_{111} (H'x) + M_{111} (H''x)$$

d'où :

$$(111) V_{111} = \Sigma (111) ix_{111} + \Sigma (211) jx_{211}/2$$

Or :

$$\begin{aligned} (111) x_{111} &= (111) x_{111} + (121) x_{121}/2 + (112) x_{112}/2 + (122) x_{122}/4 \\ (211) x_{211} &= (211) x_{211} + (221) x_{221}/2 + (212) x_{212}/2 + (222) x_{222}/4 \end{aligned}$$

ce qui donne, en tenant compte de (22), (6) et (7) :

$$\begin{aligned} (111) V_{111} &= \Sigma [PP'P'' (i + i' + i'') i + PQ'P'' (i + j' + i'') i + PP'Q'' (i + i' + j'') i \\ &\quad + PQ'Q'' (i + j' + j'') i + QP'P'' (j + i' + i'') j + QQ'P'' (j + j' + i'') j \\ &\quad + QP'Q'' (j + i' + j'') j + QQ'Q'' (j + j' + j'') j] + 0 (\mu) \end{aligned}$$

soit :

$$(111) V_{111} = \Sigma \{ (P_i + Q_j) [p'' (P'i' + Q'j') + p' (P''i'' + Q''j'')] + p'p'' (P_i^2 + Q_j^2) \} + 0 (p)$$

et :

$$V_{111} = \Sigma [\mathcal{E} (\mathcal{E}' + \mathcal{E}'') + p i^2 + q j^2] + 0 (\mu)$$

Si l'on tient compte alors de la relation (20) réalisée lorsque la population est en équilibre, on trouve finalement :

$$(29) \quad V_{111} = V + 2 (\mathcal{E} \mathcal{E}' + \mathcal{E} \mathcal{E}'' + \mathcal{E}' \mathcal{E}'') + 2 Q (\mathcal{E} + \mathcal{E}' + \mathcal{E}'') + 0 (\mu)$$

En remplaçant x_{111} et V_{111} par leurs expressions (28) et (29), on trouve :

$$(30) \quad y_{111} = - 2 (\mathcal{E} \mathcal{E}' + \mathcal{E} \mathcal{E}'' + \mathcal{E}' \mathcal{E}'') + 0 (\mu)$$

Donc, d'après (6) et (23) :

$$(31) \quad (111) = pp'p'' [1 - 2\mu (\mathcal{E} \mathcal{E}' + \mathcal{E} \mathcal{E}'' + \mathcal{E}' \mathcal{E}'')] + 0 (\mu^2)$$

et l'on établirait de même les autres fréquences d'association des gènes trois à trois :

$$(31') \quad \begin{aligned} (311) &= qp'p'' [1 - 2\mu (\mathcal{I} \mathcal{E}' + \mathcal{I} \mathcal{E}'' + \mathcal{E}' \mathcal{E}'')] + 0 (\mu^2) \\ (331) &= qq'p'' [1 - 2\mu (\mathcal{I} \mathcal{I}' + \mathcal{I} \mathcal{E}'' + \mathcal{I}' \mathcal{E}'')] + 0 (\mu^2) \\ (313) &= qp'p'' [1 - 2\mu (\mathcal{I} \mathcal{E}' + \mathcal{I} \mathcal{I}'' + \mathcal{E}' \mathcal{I}'')] + 0 (\mu^2) \\ &\text{etc...} \end{aligned}$$

D'après (7), (23) et (30) on obtient :

$$(32) \quad (111) = p^2p'^2p''^2 [1 - 8\mu (\mathcal{E} \mathcal{E}' + \mathcal{E} \mathcal{E}'' + \mathcal{E}' \mathcal{E}'')] + 0 (\mu^2)$$

et l'on établirait de même les autres fréquences d'association entre phases homozygotes des facteurs trois à trois :

$$(32') \quad \begin{aligned} (311) &= q^2p'^2p''^2 [1 - 8\mu (\mathcal{I} \mathcal{E}' + \mathcal{I} \mathcal{E}'' + \mathcal{E}' \mathcal{E}'')] + 0 (\mu^2) \\ (333) &= q^3q'^2q''^2 [1 - 8\mu (\mathcal{I} \mathcal{I}' + \mathcal{I} \mathcal{I}'' + \mathcal{I}' \mathcal{I}'')] + 0 (\mu^2) \\ (313) &= q^2p'^2q''^2 [1 - 8\mu (\mathcal{I} \mathcal{E}' + \mathcal{I} \mathcal{I}'' + \mathcal{E}' \mathcal{I}'')] + 0 (\mu^2) \\ &\text{etc...} \end{aligned}$$

Les fréquences d'association entre une phase hétérozygote et deux phases homozygotes sont obtenues à partir de (24) et (32) par la relation :

$$(112) = (110) - (111) - (113)$$

d'où :

$$(112) = p^2 p'^2 \left\{ 1 - 8\mu \mathcal{E} \mathcal{E}' - p''^2 [1 - 8\mu (\mathcal{E} \mathcal{E}' + \mathcal{E}' \mathcal{E}'' + \mathcal{E} \mathcal{E}'')] \right. \\ \left. - q''^2 [1 - 8\mu (\mathcal{E} \mathcal{E}' + \mathcal{E}' \mathcal{S}'' + \mathcal{E} \mathcal{S}'')] \right\}$$

ce qui donne toutes réductions faites :

$$(33) \quad (112) = 2 p^2 p'^2 p'' q'' [1 - 8\mu \mathcal{E} \mathcal{E}' - 4\mu (\mathcal{E} + \mathcal{E}') (\mathcal{E}'' + \mathcal{S}'')] + 0 (\mu^2)$$

et l'on obtiendrait de même les expressions analogues :

$$(33') \quad \begin{aligned} (312) &= 2 q^2 p'^2 p'' q'' [1 - 8\mu \mathcal{S} \mathcal{E}' - 4\mu (\mathcal{S} + \mathcal{E}') (\mathcal{E}'' + \mathcal{S}'')] + 0 (\mu^2) \\ (121) &= 2 p^2 p' q' p''^2 [1 - 8\mu \mathcal{E} \mathcal{E}'' - 4\mu (\mathcal{E} + \mathcal{E}'') (\mathcal{E}' + \mathcal{S}')] + 0 (\mu^2) \\ (213) &= 2 p q p'^2 q''^2 [1 - 8\mu \mathcal{E}' \mathcal{S}'' - 4\mu (\mathcal{E}' + \mathcal{S}') (\mathcal{E}'' + \mathcal{S}'')] + 0 (\mu^2) \\ &\text{etc...} \end{aligned}$$

Les fréquences d'association entre une phase homozygote et deux phases hétérozygotes sont alors obtenues à partir de (25) et (33) par la relation :

$$(122) = (120) - (121) - (123)$$

d'où :

$$(122) = 2 p^2 p' q' [1 - 4\mu \mathcal{E} (\mathcal{E}' + \mathcal{S}')] - 2 p^2 p' q' p''^2 [1 - 8\mu \mathcal{E} \mathcal{E}'' - 4\mu (\mathcal{E}' + \mathcal{S}') (\mathcal{E} + \mathcal{E}'')] - 2 p^2 p' q' q''^2 [1 - 8\mu \mathcal{E} \mathcal{S}'' - 4\mu (\mathcal{E}' + \mathcal{S}') (\mathcal{E} + \mathcal{S}'')] + 0 (\mu^2)$$

ce qui donne toutes réductions faites, en tenant compte de :

$$1 - p''^2 - q''^2 = 2 p'' q''; p'' \mathcal{E}'' + q'' \mathcal{S}'' = 0; p''^2 \mathcal{E}'' + q''^2 \mathcal{S}'' = -p'' q'' (\mathcal{E}'' + \mathcal{S}'')$$

$$(34) \quad (122) = 4 p^2 p' q' p'' q'' \left\{ 1 - 2\mu [2\mathcal{E} (\mathcal{E}' + \mathcal{S}' + \mathcal{E}'' + \mathcal{S}'') + (\mathcal{E}' + \mathcal{S}') (\mathcal{E}'' + \mathcal{S}'')] \right\} + 0 (\mu^2)$$

et l'on obtiendrait de même les expressions analogues :

$$(34') \quad \begin{aligned} (322) &= 4 q^2 p' q' p'' q'' \left\{ 1 - 2\mu [2\mathcal{S} (\mathcal{E}' + \mathcal{S}' + \mathcal{E}'' + \mathcal{S}'') + (\mathcal{E}' + \mathcal{S}') (\mathcal{E}'' + \mathcal{S}'')] \right\} + 0 (\mu^2) \\ (212) &= 4 p q p'^2 p'' q'' \left\{ 1 - 2\mu [2\mathcal{E}' (\mathcal{E} + \mathcal{S} + \mathcal{E}'' + \mathcal{S}'') + (\mathcal{E} + \mathcal{S}) (\mathcal{E}'' + \mathcal{S}'')] \right\} + 0 (\mu^2) \\ (223) &= 4 p q p' q' q''^2 \left\{ 1 - 2\mu [2\mathcal{S}'' (\mathcal{E} + \mathcal{S} + \mathcal{E}' + \mathcal{S}') + (\mathcal{E} + \mathcal{S}) (\mathcal{E}' + \mathcal{S}')] \right\} + 0 (\mu^2) \\ &\text{etc...} \end{aligned}$$

Enfin la probabilité d'association des trois phases hétérozygotes est obtenue à partir de (26) et (34) par la relation :

$$(222) = (220) - (221) - (223)$$

d'où :

$$(222) = 4 p q p' q' \left\{ 1 - 2 \mu (\mathcal{E} + \mathcal{S}) (\mathcal{E}' + \mathcal{S}') \right. \\ \left. - p''^2 \left[1 - 2 \mu [2 \mathcal{E}'' (\mathcal{E}' + \mathcal{S}' + \mathcal{E} + \mathcal{S}) + (\mathcal{E} + \mathcal{S}) (\mathcal{E}' + \mathcal{S}')] \right] \right. \\ \left. - q''^2 \left[1 - 2 \mu [2 \mathcal{S}'' (\mathcal{E}' + \mathcal{S}' + \mathcal{E} + \mathcal{S}) + (\mathcal{E} + \mathcal{S}) (\mathcal{E}' + \mathcal{S}')] \right] \right\}$$

ce qui donne toutes réductions faites :

$$(35) \quad (222) = \frac{8 p q p' q' p'' q''}{\left\{ 1 - 2 \mu [(\mathcal{E} + \mathcal{S})(\mathcal{E}' + \mathcal{S}') + (\mathcal{E}' + \mathcal{S}')(\mathcal{E}'' + \mathcal{S}'') + (\mathcal{E}'' + \mathcal{S}'')(\mathcal{E} + \mathcal{S})] \right\} + 0 (\mu^2)}$$

Les formules (32), (33), (34), (35) permettent d'écrire immédiatement les coefficients f_{ijk} de Fisher pour les probabilités d'association des phases des facteurs trois à trois :

$$(36) \quad \begin{aligned} f_{111} &= -8 \mu (\mathcal{E} \mathcal{E}' + \mathcal{E}' \mathcal{E}'' + \mathcal{E}'' \mathcal{E}) \\ f_{113} &= -8 \mu (\mathcal{E} \mathcal{E}' + \mathcal{E}' \mathcal{S}'' + \mathcal{S}'' \mathcal{E}) \\ &\dots \dots \dots \\ f_{112} &= -4 \mu [2 \mathcal{E} \mathcal{E}' + (\mathcal{E} + \mathcal{E}') (\mathcal{E}'' + \mathcal{S}'')] \\ f_{132} &= -4 \mu [2 \mathcal{E} \mathcal{S}' + (\mathcal{E} + \mathcal{S}') (\mathcal{E}'' + \mathcal{S}'')] \\ &\dots \dots \dots \\ f_{122} &= -2 \mu [2 \mathcal{E} (\mathcal{E}' + \mathcal{S}' + \mathcal{E}'' + \mathcal{S}'') + (\mathcal{E}' + \mathcal{S}') (\mathcal{E}'' + \mathcal{S}'')] \\ f_{322} &= -2 \mu [2 \mathcal{S}' (\mathcal{E}' + \mathcal{S}' + \mathcal{E}'' + \mathcal{S}'') + (\mathcal{E}' + \mathcal{S}') (\mathcal{E}'' + \mathcal{S}'')] \\ &\dots \dots \dots \\ f_{222} &= -2 \mu [(\mathcal{E} + \mathcal{S})(\mathcal{E}' + \mathcal{S}') + (\mathcal{E}' + \mathcal{S}') (\mathcal{E}'' + \mathcal{S}'') + (\mathcal{E}'' + \mathcal{S}'') (\mathcal{E} + \mathcal{S})] \end{aligned}$$

CHAPITRE V

CORRELATIONS RELATIVES AU CARACTÈRE PRIMAIRE DANS UNE POPULATION EN EQUILIBRE

Notations. — Nous conserverons dans ce chapitre où nous supposons encore la population en équilibre, les notations du chapitre précédent. Mais nous supposerons de plus que la contribution H du facteur (H) au caractère primaire se décompose en deux parties :

— L'une, que nous appellerons « contribution génétique » G, est la somme de deux aléatoires du deuxième ordre $\mathcal{H}$ et $\mathcal{H}_1$ correspondant respectivement aux contributions additives des deux gènes caractérisant le facteur (H) ;

— L'autre, est un résidu d qui traduit l'effet de dominance.

Nous poserons donc :

$$G = \mathcal{H} + \mathcal{H}_1; \quad H = G + d = \mathcal{H} + \mathcal{H}_1 + d$$

d'où :

$$x = \Sigma H = \Sigma G + \Sigma d$$

et finalement :

$$x = z + D$$

en posant :

$$z = \Sigma G \quad \text{et} \quad D = \Sigma d$$

Nous conviendrons de désigner les quantités relatives à G par les mêmes lettres que les quantités analogues relatives à H, mais affectées de l'indice 0. Si l'on représente alors par t et s les valeurs que peuvent prendre les aléatoires $\mathcal{H}$ et $\mathcal{H}_1$, l'aléatoire du troisième ordre G peut prendre les trois valeurs :

$$i_0 = 2t; \quad j_0 = t + s; \quad k_0 = 2s$$

auxquelles correspondent pour l'aléatoire du troisième ordre d les trois valeurs que nous désignerons par :

$$m = i - i_0; \quad n = j - j_0; \quad r = k - k_0$$

*
* *

Détermination du rôle de la dominance. - Ajustements par moindres carrés.

Premier mode d'ajustement. — Considérons séparément chaque facteur et déterminons pour celui-ci la contribution génétique G de telle sorte que $M(d^2)$ soit minimum : on obtient en annulant les dérivées partielles par rapport à s et t de l'expression :

$$M(d^2) = P(i - i_0)^2 + 2Q(j - j_0)^2 + R(k - k_0)^2$$

les deux conditions :

$$\begin{cases} P(i-2t) + Q(j-s-t) = 0 \\ Q(j-s-t) + R(k-2s) = 0 \end{cases}$$

soit, avec l'approximation adoptée :

(1)

$$\begin{cases} pi + qj = pio + qjo \\ pj + qk = pjo + qko \end{cases}$$

conditions évidemment identiques à celles que l'on a dans le cas du random mating (1), puisqu'il s'agit de conséquences de la loi de Hardy.

Et, en tenant compte de la relation $2j_0 = i_0 + k_0$, le système (1) admet la seule solution

(2)

$$i_0 = \frac{p^2 i - q^2 k}{p}; \quad k_0 = \frac{q^2 k - p^2 i}{q}$$

Deuxième mode d'ajustement. — Considérons l'ensemble des facteurs déterminant le caractère primaire et déterminons la contribution génétique de chacun des facteurs de telle sorte que la quantité $M(D^2)$ soit minimum. Or on a :

$$\begin{aligned} M(D^2) = & (111 \dots) (i + i' + \dots - 2t - 2t' - \dots)^2 \\ & + (211 \dots) (j + i' + \dots - t - s - 2t' - \dots)^2 \\ & + (311 \dots) (k + i' + \dots - 2s - 2t' - \dots)^2 \\ & + (121 \dots) (i + j' + \dots - 2t - t' - s' - \dots)^2 \\ & + (221 \dots) (j + j' + \dots - t - s - t' - s' - \dots)^2 \\ & + (321 \dots) (k + j' + \dots - 2s - t' - s' - \dots)^2 \\ & + \dots \end{aligned}$$

et, en annulant les dérivées partielles de $M(D^2)$ par rapport à t , à s , à t' , à s' ,... on obtient pour chaque facteur deux équations ; par exemple :

$$\begin{aligned} -\frac{1}{2} \frac{\partial M(D^2)}{\partial t} = & 2(111 \dots) (x_{111} - z_{111}) + (211 \dots) (x_{211} - z_{211}) \\ & + 2(121 \dots) (x_{121} - z_{121}) + (221 \dots) (x_{221} - z_{221}) \\ & + \dots = 0 \end{aligned}$$

ce qui s'écrit :

$$2(100 \dots) (x_{100} - z_{100}) + (200 \dots) (x_{200} - z_{200}) = 0$$

(1) Cf. G. MALÉCOT, Thèse, Paris 1939, p. 45, (2°)

ou enfin :

$$(3) \quad (100) (x_{100} - z_{100}) = 0$$

Les équations qui définissent l'ajustement sont donc :

$$(3') \quad \begin{aligned} PI + QJ &= PI_0 + QJ_0 \\ QJ + RK &= QJ_0 + RK_0 \\ P'I' + Q'J' &= P'I'_0 + Q'J'_0 \\ Q'J' + R'K' &= Q'J'_0 + R'K'_0 \\ &\text{etc...} \end{aligned}$$

En additionnant membre à membre ces équations deux à deux, on trouve les relations :

$$(4) \quad \begin{aligned} PI + 2 QJ + RK &= PI_0 + 2 QJ_0 + RK_0 = 0 \\ P'I' + 2 Q'J' + R'K' &= P'I'_0 + 2 Q'J'_0 + R'K'_0 = 0 \\ &\text{etc...} \end{aligned}$$

ou encore en tenant compte de la loi de Hardy :

$$(4') \quad \begin{aligned} p^2 I_0 + 2 pq J_0 + q^2 K_0 &= 0 \\ p'^2 I'_0 + 2 p' q' J'_0 + q'^2 K'_0 &= 0 \\ &\text{etc...} \end{aligned}$$

Or, en remplaçant $I_0, J_0, K_0...$ par leurs expressions (cf. p. 45) :

$$I_0 = i_0 + \Sigma' [P' (1 + f_{11}) i'_0 + 2 Q' (1 + f_{12}) j'_0 + R' (1 + f_{13}) k'_0] \\ \text{etc...}$$

où les coefficients f sont les coefficients de Fisher d'association des phases des facteurs deux à deux, et où la somme Σ' est étendue à tous les facteurs sauf (H), et en remarquant que :

$$PP' (1 + f_{11}) + 2 QP' (1 + f_{21}) + RP' (1 + f_{31}) = P' = p'^2 \\ \text{etc...}$$

les équations (4') donnent la relation :

$$(5) \quad \Sigma (p^2 i_0 + 2 pq j_0 + q^2 k_0) = 0$$

ou encore en tenant compte de $j_0 = (i_0 + k_0)/2$:

$$(5') \quad \Sigma (p i_0 + q k_0) = 0$$

Equivalence des deux ajustements. — Remarquons que la solution (2) du système (1) vérifie les relations (5) et (5') conséquences du système (3'). Montrons que cette solution vérifie avec l'approximation adoptée les équations (3').

En comparant les expressions de I, J, K et de I₀, J₀, K₀, on a :

$$\begin{aligned} P(I - I_0) &= Pm + \Sigma' [(11) m' + (12) n' + (13) r'] \\ 2 Q(J - J_0) &= 2 Qn + \Sigma' [(21) m' + (22) n' + (23) r'] \\ &\text{etc...} \end{aligned}$$

La première équation (3') s'écrit alors :

$$(6) \quad Pm + Qn + \Sigma' [(11) m' + (12) n' + (13) r'] = 0$$

en posant comme d'habitude :

$$(11) = (11) + (21)/2$$

ou, en utilisant les expressions (IV-25) établies au chapitre précédent et en tenant compte de : $p\mathcal{E} + q\mathcal{S} = 0$,

$$\begin{cases} (11) = pp'^2 (1 - 4\mu\mathcal{E}) \\ (13) = pq'^2 (1 - 4\mu\mathcal{S}) \\ (12) = 2pp'q' [1 - 2\mu(\mathcal{E}' + \mathcal{S}')] \end{cases}$$

l'équation (6) s'écrit :

$$\begin{aligned} p^2m + pqn + p\Sigma' \{ p'^2m' + 2p'q'n' + q'^2r' - 4\mu\mathcal{E} \\ [p'^2\mathcal{E}'m' + p'q'(\mathcal{E}' + \mathcal{S}')n' + q'^2\mathcal{S}'r'] \} = 0 \end{aligned}$$

Or, d'après (5) et (IV-10) on a :

$$\Sigma (p^2m + 2pqn + q^2r) = p^2m + 2pqn + q^2r + \Sigma' (p'^2m' + 2p'q'n' + q'^2r') = 0$$

et finalement, en tenant compte de $p'\mathcal{E}' + q'\mathcal{S}' = 0$, la première équation (3') devient :

$$(6') \quad pm + qn = p^2m + 2pqn + q^2r + 4\mu\mathcal{E}\Sigma' \{ p'\mathcal{E}' [p'(m' - n') + q'(n' - r')] \}$$

La deuxième équation (3') s'écrit de même :

$$(6'') \quad pn + qr = p^2m + 2pqn + q^2r + 4\mu\mathcal{S}\Sigma' \{ p'\mathcal{E}' [p'(m' - n') + q'(n' - r')] \}$$

Mais les solutions (2) du système (1) satisfont aux relations suivantes que l'on déduit immédiatement de (1) :

$$\begin{aligned} pm + qn &= 0; & pn + qr &= 0; & p^2m + 2pqn + q^2r &= 0; \\ p'(m' - n') + q'(n' - r') &= 0 \end{aligned}$$

Elles vérifient donc les équations (6') et (6''); ce sont par conséquent aussi des solutions du système (3').

Nous voyons qu'avec l'approximation convenue, les deux modes d'ajustement ont la même solution (2) : ils sont équivalents, et nous pourrons simultanément utiliser les conséquences que l'on peut déduire de l'un ou de l'autre.

Interprétation des deux modes d'ajustement. — Les équations (1) qui définissent le premier ajustement signifient que lorsque l'un des gènes concernant le facteur (H) est déterminé ($\mathcal{H}$ ou $\mathcal{H}_1$ fixé à la valeur t ou s), on a :

$$M(d) = 0$$

Nous appellerons cette propriété la *propriété (A)*.

De même les équations (3') qui définissent le deuxième ajustement signifient que si un gène de l'un quelconque des facteurs est déterminé, on a :

$$M(D) = 0$$

Propriété que nous appellerons la *propriété (B)*.

L'équivalence des deux ajustements nous permettra d'utiliser les propriétés (A) et (B) en donnant aux contributions génétiques des différents facteurs les valeurs (2).

*
* *

Variances relatives à chaque facteur : $M(H^2)$; $M(G^2)$; $M(d^2)$.

On a :

$$M(H^2) = M[(G + d)^2] = M(G^2) + M(d^2) + 2 M(Gd)$$

Or :

$$M(Gd) = M(\mathcal{H}d) + M(\mathcal{H}_1d) = 0 \text{ d'après la propriété (A) ;}$$

donc :

$$(7) \quad M(H^2) = M(G^2) + M(d^2)$$

Proposons-nous de calculer les variances partielles $M(G^2)$ et $M(d^2)$ sachant que l'on a :

$$(8) \quad M(H^2) = p^2 i_0^2 + 2 pq j_0^2 + q^2 k_0^2$$

La variance de la contribution génétique G s'écrit :

$$M(G^2) = p^2 i_0^2 + 2 pq j_0^2 + q^2 k_0^2$$

ou, en tenant compte de $j_0 = (i_0 + k_0)/2$:

$$M(G^2) = p(p + \frac{q}{2}) i_0^2 + pq i_0 k_0 + q(q + \frac{p}{2}) k_0^2$$

et en donnant à i_0 et k_0 les valeurs (2) on trouve finalement :

$$(9) \quad M(G^2) = (p^2i - q^2k)^2/2pq$$

On a de même :

$$M(d^2) = p^2m^2 + 2qpn^2 + q^2r^2$$

Or on déduit immédiatement de (2) :

$$\begin{cases} m = i - (p^2i - q^2k)/p = (pi + qk)q/p \\ r = k - (q^2k - p^2i)/q = (pi + qk)p/q \\ n = -(p^2m + q^2r)/2pq = -(pi + qk) \end{cases}$$

d'où, toutes réductions faites :

$$(10) \quad M(d^2) = (pi + qk)^2$$

Nous obtenons évidemment les mêmes résultats que dans le cas du random mating, puisque les valeurs (2) des contributions génétiques sont les mêmes et puisque la loi de Hardy est vérifiée.

*
* *

Variance totale : $V = M(x^2)$.

On a :

$$M(x^2) = M[(z + D)^2] = M(z^2) + M(D^2) + 2M(zD)$$

Or il résulte de la propriété (B) que :

$$M(zD) = M(\mathcal{H}D) + M(\mathcal{H}_1D) + M(\mathcal{H}'D) + M(\mathcal{H}'_1D) + \dots = 0$$

puisque chacun des termes de la somme figurant au deuxième membre est nul. Donc :

$$(11) \quad M(x^2) = M(z^2) + M(D^2)$$

Calcul de la variance génétique totale : $V_0 = M(z^2)$.

Nous pouvons prévoir que la variance génétique V_0 sera différente de celle que l'on a dans le cas du random mating, car, le mode d'union adopté introduisant une liaison entre les différents facteurs, on a :

$$M(z^2) \neq \sum M(G^2)$$

Nous remarquons que l'on peut écrire la variance génétique totale :

$$M(z^2) = M[z(x - D)] = M(zx) - M(zD) = M(zx)$$

donc :

$$M(z^2) = \Sigma M(Gx) = \Sigma (p^2 i_0 I + 2 p q j_0 J + q^2 k_0 K)$$

et en tenant compte de $j_0 = (i_0 + k_0)/2$:

$$M(z^2) = \Sigma [p i_0 (p I + q J) + q k_0 (p J + q K)]$$

ou, d'après (3') :

$$M(z^2) = \Sigma [p i_0 (p I_0 + q J_0) + q k_0 (p J_0 + q K_0)]$$

et finalement :

$$(12) \quad \boxed{M(z^2) = \Sigma v_0}$$

en posant :

$$v_0 = p^2 i_0 I^2 + 2 p q j_0 J_0 + q^2 k_0 K_0$$

Remarquons que l'on a :

$$v_0 = p i_0 (p I + q J) + q k_0 (p J + q K) = p i_0 \mathcal{E} + q k_0 \mathcal{S}$$

ou :

$$(13) \quad \boxed{v_0 = p \mathcal{E} (i_0 - k_0)}$$

Or, les expressions de I_0, J_0, K_0 données page 59 permettent d'écrire :

$$p^2 I_0 + p q J_0 = p^2 i_0 + p q j_0 + \Sigma' [(11) i_0' + (13) k_0']$$

car :

$$\begin{aligned} PP' (1 + f_{11}) i_0' + 2 PQ' (1 + f_{12}) j_0' + PR' (1 + f_{13}) k_0' &= (11) i_0' + (12) j_0' + (13) k_0' \\ &= (11) i_0' + (12) i_0'/2 + (12) k_0'/2 + (13) k_0' \\ &= (11) i_0' + (13) k_0' \end{aligned}$$

et de même :

$$QP' (1 + f_{21}) i_0' + 2 QQ' (1 + f_{22}) j_0' + QR' (1 + f_{23}) k_0' = (21) i_0'/2 + (23) k_0'/2$$

En remplaçant alors les fréquences d'association (11) et (13) par leurs expressions (IV-24) établies au chapitre précédent, on obtient :

$$\begin{aligned} p^2 I_0 + p q J_0 &= p^2 i_0 + p q j_0 + \Sigma' [p p' (1 - 2 \mu \mathcal{E} \mathcal{E}') i_0' + p q' (1 - 2 \mu \mathcal{E} \mathcal{S}') k_0'] \\ &= p^2 i_0 + p q j_0 + p \Sigma' (p' i_0' + q' k_0') - 2 \mu p \mathcal{E} \Sigma' (p' \mathcal{E}' i_0' + q' \mathcal{S}' k_0') \end{aligned}$$

Ou encore, en tenant compte de : $p \mathcal{E} + q \mathcal{S} = 0$, et en remarquant que d'après (5') on a :

$$\Sigma' (p' i_0' + q' k_0') = - (p i_0 + q k_0)$$

on peut écrire :

$$p^2 I_0 + p q J_0 = p^2 i_0 + p q j_0 + 2 \mu p^2 \mathcal{E}^2 (i_0 - k_0) - p (p i_0 + q k_0) - 2 \mu p \mathcal{E} \Sigma p \mathcal{E}' (i_0 - k_0)$$

On établirait de même :

$$pqJ_o + q^2K_o = pqj_o + q^2k_o + 2 \mu q^2 \mathcal{S}^2 (k_o - i_o) - q (pi_o + qk_o) - 2 \mu q \mathcal{S} \Sigma q \mathcal{S} (k_o - i_o)$$

ou, en tenant compte de $p\mathcal{E} = -q\mathcal{S}$:

$$pqJ_o + q^2K_o = pqj_o + q^2k_o - 2 \mu p^2 \mathcal{E}^2 (i_o - k_o) - q (pi_o + qk_o) + 2 \mu p \mathcal{E} \Sigma p \mathcal{E} (i_o - k_o)$$

et si l'on écrit l'équation (13) :

$$v_o = p\mathcal{E} i_o - p\mathcal{E} k_o = p\mathcal{E} i_o + q\mathcal{S} k_o = (p^2I_o + pqJ_o) i_o + (pqJ_o + q^2K_o) k_o$$

on obtient :

$$v_o = p^2i_o^2 + 2 pqj_o^2 + q^2k_o^2 + 2 \mu p^2 \mathcal{E}^2 (i_o - k_o)^2 - (pi_o + qk_o)^2 - 2 \mu p \mathcal{E} (i_o - k_o) \Sigma p \mathcal{E} (i_o - k_o)$$

ce qui donne finalement d'après (13) :

$$(14) \quad v_o = M (G^2) + 2 \mu v_o^2 - 2 \mu v_o \Sigma v_o - (pi_o + qk_o)^2$$

Enfin, comme d'après (2) $pi_o + qk_o = 0$, l'expression de la variance génétique totale est :

$$(15) \quad V_o = \Sigma M (G^2) + 2 \mu \Sigma v_o^2 - 2 \mu (\Sigma v_o)^2$$

En particulier, si l'on suppose que le nombre n des facteurs déterminant le caractère primaire est assez grand, et que les effets de chaque facteur sont comparables, v_o est de l'ordre de V_o/n , Σv_o^2 de l'ordre de V_o^2/n , donc Σv_o^2 est négligeable devant $(\Sigma v_o)^2 = V_o^2$. On a alors la valeur approchée suivante de la variance génétique totale :

$$(15'') \quad V_o = \Sigma M (G^2) - 2 \mu V_o^2$$

Calcul de $V = M (x^2)$.

La variance totale V peut s'écrire :

$$V = M (x^2) = \Sigma M (Hx) = \Sigma (p^2iI + 2 pqjJ + q^2kK)$$

donc :

$$(16) \quad v = p^2iI + 2 pqjJ + q^2kK$$

en posant :

Comme on a par hypothèse :

$$2 pqj = - (p^2I + q^2K)$$

v peut s'écrire :

$$(16') \quad v = p^2I (i - j) + q^2K (k - j)$$

Or, les expressions de I et K :

$$\begin{cases} I = i + \Sigma' [P' (1 + f_{11}) i' + 2 Q' (1 + f_{12}) j' + R' (1 + f_{13}) k'] \\ K = k + \Sigma' [P' (1 + f_{31}) i' + 2 Q' (1 + f_{32}) j' + R' (1 + f_{33}) k'] \end{cases}$$

deviennent si l'on tient compte des relations :

$$P'i' + 2 Q'j' + R'k' = 0$$

et :

$$PP' (1 + f_{11}) + 2 PQ' (1 + f_{12}) + PR' (1 + f_{13}) = P$$

qui entraîne :

$$P'f_{11} + 2 Q'f_{12} + R'f_{13} = 0$$

etc...

$$\begin{cases} I = i + \Sigma' [P' (i' - j') f_{11} - R' (j' - k') f_{13}] \\ K = k + \Sigma' [P' (i' - j') f_{31} - R' (j' - k') f_{33}] \end{cases}$$

ou, en donnant aux coefficients f les valeurs (IV-27) :

$$\begin{cases} I = i - 8 \mu \mathcal{E} \Sigma' p' \mathcal{E}' [p' (i' - j') + q' (j' - k')] \\ K = k - 8 \mu \mathcal{S} \Sigma' p' \mathcal{E}' [p' (i' - j') + q' (j' - k')] \end{cases}$$

Or, en tenant compte de $2 j_0 = i_0 + k_0$, le système (1) donne :

$$(17) \quad p (i - j) + q (j - k) = p (i_0 - j_0) + q (j_0 - k_0) = (i_0 - k_0)/2$$

Donc :

$$\begin{cases} I = i - 8 \mu \mathcal{E} \Sigma' p' \mathcal{E}' (i_0' - k_0')/2 \\ K = k - 8 \mu \mathcal{S} \Sigma' p' \mathcal{E}' (i_0' - k_0')/2 \end{cases}$$

Or, d'après (13) :

$$\Sigma p \mathcal{E} (i_0 - k_0)/2 = \Sigma v_0/2 = V_0/2$$

donc :

$$\Sigma' p' \mathcal{E}' (i_0' - k_0')/2 = (V_0 - v_0)/2$$

et finalement :

$$(18) \quad \begin{cases} I = i - 4 \mu \mathcal{E} (V_0 - v_0) \\ K = k - 4 \mu \mathcal{S} (V_0 - v_0) \end{cases}$$

Remplaçons dans (16') I et K par les valeurs (18) :

$$v = p^2 i (i - j) + q^2 k (k - j) - 4 \mu (V_0 - v_0) [p^2 \mathcal{E} (i - j) + q^2 \mathcal{S} (k - j)]$$

ce qui s'écrit en tenant compte de (17) et des relations :

$$\begin{aligned} p^2 i + q^2 k &= -2 p q j & \text{et} & & p \mathcal{E} + q \mathcal{F} &= 0 \\ v &= p^2 i^2 + 2 p q j^2 + q^2 k^2 - 2 \mu (V_0 - v_0) p \mathcal{E} (i_0 - k_0) \end{aligned}$$

et enfin d'après (13) :

$$(19) \quad v = M(H^2) - 2 \mu V_0 v_0 + 2 \mu v_0^2$$

La variance totale $V = \Sigma v$ a alors comme expression :

$$(20) \quad V = M(x^2) = \Sigma M(H^2) - 2 \mu V_0 v_0 + 2 \mu \Sigma v_0^2$$

Ici encore, si l'on suppose le nombre n des facteurs dont dépend le caractère primaire assez grand, le terme $\mu \Sigma v_0^2$ est négligeable et l'on a la valeur approchée suivante de la variance totale :

$$(20') \quad V = \Sigma M(H^2) - 2 \mu V_0 v_0$$

Calcul de $M(D^2)$:

Il résulte de (11) et des expressions (15) et (20) des variances V_0 et V que l'on a :

$$M(D^2) = \Sigma M(H^2) - \Sigma M(G^2)$$

et donc d'après (7) :

$$(21) \quad M(D^2) = \Sigma M(d^2)$$

Remarque. — Les résultats que nous venons d'établir pour les variances totales diffèrent de ceux que l'on aurait obtenus dans le cas du random mating, mais ils sont en tous points analogues à ceux qu'a trouvés M. MALÉCOT dans le cas d'homogamie qu'il a étudié ⁽¹⁾.

*
* *

Corrélation entre conjoints :

Nous désignerons par des lettres surlignées $\bar{H}$, $\bar{x}$, $\bar{z}$... les aléatoires relatives au conjoint de l'individu considéré.

On a :

$$\begin{aligned} M(x\bar{x}) &= M[(H + H' + H'' + \dots)(\bar{H} + \bar{H}' + \bar{H}'' + \dots)] \\ &= \Sigma \Sigma' M(H\bar{H}) + \Sigma M(H\bar{H}) \end{aligned}$$

(1) Cf. G. MALÉCOT, *Thèse*, Paris, 1939, p. 84 et suivantes.

Nous sommes donc amenés à calculer $M(H\bar{H}')$ et $M(H\bar{H})$. Or :

$$M(H\bar{H}') = p^2 p'^2 i i' (1 + \mu y_{10} + \mu y_{01}) + 2 p q p' j i' (1 + \mu y_{20} + \mu y_{01}) + q^2 p'^2 k i' (1 + \mu y_{30} + \mu y_{01}) \\ + 2 p^2 p' q' i j' (1 + \mu y_{10} + \mu y_{02}) + 4 p q p' q' j j' (1 + \mu y_{20} + \mu y_{02}) + 2 q^2 p' q' k j' (1 + \mu y_{30} + \mu y_{02}) \\ + p^2 q'^2 i k' (1 + \mu y_{10} + \mu y_{03}) + 2 p q q'^2 j k' (1 + \mu y_{20} + \mu y_{03}) + q^2 q'^2 k k' (1 + \mu y_{30} + \mu y_{03})$$

ou :

$$M(H\bar{H}') = (p^2 i + 2 p q j + q^2 k) (p'^2 i' + 2 p' q' j' + q'^2 k') \\ + \mu (p^2 i + 2 p q j + q^2 k) (p'^2 i' y_{01} + 2 p' q' j' y_{02} + q'^2 k' y_{03}) \\ + \mu (p'^2 i' + 2 p' q' j' + q'^2 k') (p^2 i y_{10} + 2 p q j y_{20} + q^2 k y_{30})$$

On a donc, par suite des relations : $p^2 i + 2 p q j + q^2 k = 0$:

$$M(H\bar{H}') = 0$$

On écrit de même :

$$M(H\bar{H}) = p^4 i^2 (1 + 2 \mu y_{100}) + 2 p^3 q i j (1 + \mu y_{100} + \mu y_{200}) + p^2 q^2 i k (1 + \mu y_{100} + \mu y_{300}) \\ + 2 p^3 q i j (1 + \mu y_{100} + \mu y_{200}) + 4 p^2 q^2 j^2 (1 + 2 \mu y_{200}) + 2 p q^2 j k (1 + \mu y_{200} + \mu y_{300}) \\ + p^2 q^2 i k (1 + \mu y_{100} + \mu y_{300}) + 2 p q^2 j k (1 + \mu y_{200} + \mu y_{300}) + q^4 k^2 (1 + 2 \mu y_{300})$$

ou :

$$M(H\bar{H}) = (p^2 i + 2 p q j + q^2 k)^2 + 2 \mu (p^2 i + 2 p q j + q^2 k) (p^2 i y_{100} + 2 p q j y_{200} + q^2 k y_{300})$$

et :

$$M(H\bar{H}) = 0$$

On a par conséquent :

$$(22) \quad \boxed{M(x\bar{x}) = 0}$$

Le coefficient de corrélation entre époux est nul : le mode d'union adopté ne présente pas le caractère d'homogamie.

Remarque. — On déduit des relations (1) les égalités :

$$p^2 i_0 + 2 p q j_0 + q^2 k_0 = 0$$

Un raisonnement absolument analogue au précédent montrerait donc que les aléatoires z et $\bar{z}$ sont indépendantes : les expressions de $M(G\bar{G}')$ et de $M(G\bar{G})$ en fonction des contributions génétiques $i_0, j_0, k_0...$ donnent en effet :

$$M(G\bar{G}') = 0 \quad \text{et} \quad M(G\bar{G}) = 0$$

et par conséquent :

$$(22') \quad \boxed{M(z\bar{z}) = 0}$$

*
* *

Corrélation Père - Fils :

Nous désignerons par les lettres affectées de l'indice 1 : $x_1, z_1, H_1, \dots$ les aléatoires relatives au fils.

Si l'on suppose fixées les constitutions génétiques des deux parents, M. MALÉCOT a montré ⁽¹⁾ que la valeur moyenne de la contribution *génétique* du facteur (H) chez l'enfant est donnée en fonction des contributions génétiques G et $\bar{G}$ du même facteur chez les parents, par l'expression :

$$M(G_1) = (G + \bar{G})/2$$

d'où :

$$(23) \quad M(z_1) = (z + \bar{z})/2$$

Or, on a :

$$(24) \quad M(xx_1) = M(xz_1) + M(xD_1)$$

Nous sommes donc amenés à calculer $M(xz_1)$ et $M(xD_1)$.

Calcul de $M(xz_1)$:

Si l'on remarque que, lorsque x est fixé, on a d'après (23) :

$$M(xz_1) = xM(z_1) = xz/2 + xM(\bar{z})/2$$

on peut écrire d'une façon générale :

$$M(xz_1) = M(xz)/2 + M(x\bar{z})/2$$

Or le deuxième terme $M(x\bar{z})/2$ est nul puisque le coefficient de corrélation entre conjoints est nul ; par ailleurs, on a :

$$M(xz) = M[(z + D)z] = M(z^2) + M(zD)$$

et $M(zD) = 0$ d'après la propriété (B). Donc finalement :

$$(25) \quad M(xz_1) = V_0/2$$

Remarque. — Un raisonnement analogue montrerait que :

$$(25') \quad M(zz_1) = V_0/2$$

(1) Cf. G. MALÉCOT, *Thèse*, Paris, 1939, p. 32 et 88.

Calcul de $M(xD_1)$:

On a :

$$M(xD_1) = \sum M(xd_1)$$

Or :

$$M(xd_1) = M(Hd_1) + M(H'd_1) + M(H''d_1) + \dots$$

Donc :

$$M(xD_1) = \sum \sum M(Hd'_1) + \sum M(Hd_1)$$

Calculons $M(Hd_1)$ et $M(Hd'_1)$.

La probabilité pour que $H = i$ et que $d_1 = m$ (c'est-à-dire que $H_1 = i$) est :

$$(100) * (100) + (100) * (200)/2 = (100) * (100) = p^2 * p$$

De même, la probabilité pour que $H = i$ et que $d_1 = n$ (c'est-à-dire que $H_1 = j$) est :

$$(100) * (300) + (100) * (200)/2 = (100) * (300) = p^2 * q$$

La probabilité pour que $H = j$ et que $d_1 = n$ (c'est-à-dire que $H_1 = j$) est :

$$(200) * (100)/2 + (200) * (300)/2 + (200) * (200)/2 = (200) * (000)/2 = pq * (000)$$

en désignant toujours par $\pi * (000)$ la probabilité d'union d'un individu de probabilité π avec un individu quelconque de la population.

Et ainsi de suite... Nous pouvons établir le tableau suivant qui donne les probabilités d'association des différentes valeurs possibles des aléatoires H et d_1 .

TABLEAU I

$H \backslash d_1$	m	n	r
i	$p^2 * p$	$p^2 * q$	0
j	$pq * p$	$pq * (000)$	$pq * q$
k	0	$q^2 * p$	$q^2 * q$

De même, la probabilité pour que $H = i$ et $d'_1 = m'$ (c'est-à-dire que $H'_1 = i'$) est :

$$(11) * (01) + (12) * (01)/2 + (11) * (02)/2 + (12) * (02)/4 = (11) * p'$$

La probabilité pour que $H = j$ et que $d'_1 = n'$ (c'est-à-dire que $H'_1 = j'$) est :

$$\begin{aligned} & (21) * (02)/2 + (21) * (03) + (22) * (01)/2 + (22) * (02)/2 \\ & + (22) * (03)/2 + (23) * (01) + (23) * (02)/2 \\ & = (23) * (01) + (23) * (02)/2 + (21) * (03) + (21) * (02)/2 \\ & = (23) * p' + (21) * q' \end{aligned}$$

Et ainsi de suite... Nous pouvons dresser le tableau II suivant, qui donne les probabilités d'association des différentes valeurs possibles des aléatoires H et d'_1 :

TABLEAU II

H \ d'_1	m'	n'	r'
i	$(11) * p'$	$(11) * q' + (13) * p'$	$(13) * q'$
j	$(21) * p'$	$(21) * q' + (23) * p'$	$(23) * q'$
k	$(31) * p'$	$(31) * q' + (33) * p'$	$(33) * q'$

Le tableau I donne immédiatement :

$$M(Hd_1) = (p^2i + pqj) * (pm + qn) + (pqj + q^2k) * (pn + qr)$$

ce qui s'écrit, en développant les produits symboliques et en remarquant que les équations (1) s'écrivent :

$$pm + qn = 0 (\mu^2) ; \quad pn + qr = 0 (\mu^2)$$

$$M(Hd_1) = \mu [(p^2i + pqj) (pmT + qnS) + (pqj + q^2k) (pnT + qrS)] + 0 (\mu^2)$$

et finalement, puisque :

$$T = 0 (\mu) \quad \text{et} \quad S = 0 (\mu)$$

on a :

$$M(Hd_1) = 0 (\mu^2)$$

Le tableau II donne de la même façon :

$$M(Hd'_1) = [(11)i + (21)j + (31)k] * (p'm' + q'n') + [(13)i + (23)j + (33)k] * (p'n' + q'r')$$

ce qui s'écrit en remarquant que :

$$(11)i + (21)j + (31)k = p' (p^2i + 2.pqj + q^2k) + 0 (\mu) = 0 (\mu)$$

et en tenant compte de :

$$p'm' + q'n' = 0 (\mu^2) \quad \text{etc...}$$

$$M(Hd'_1) = 0 (\mu^3)$$

Si l'on remarque enfin que chacun des termes qui figure dans le développement de $M(Hd_1)$ ou de $M(Hd'_1)$ est de l'ordre de V_0/n , n étant le nombre de facteurs, et donc que le coefficient de μ^2 dans $M(Hd_1)$ et de μ^3 dans $M(Hd'_1)$ est de l'ordre de V_0/n , on voit que :

$$\Sigma \Sigma Hd'_1 = V_0 \cdot 0 (n\mu^3) \quad \text{et} \quad \Sigma Hd_1 = V_0 \cdot 0 (\mu^2)$$

et finalement :

$$M(xD_1) = V_0 \cdot [0 (\mu^2) + 0 (n\mu^3)]$$

Donc, les équations (24) et (25) donnent, avec l'approximation que nous avons adoptée :

$$(26) \quad M(x x_1) = V_0/2$$

Le coefficient de corrélation père-fils a alors pour expression :

$$(27) \quad \rho = \frac{V_0}{2V}$$

* * *

Comparaison avec le cas du random mating :

L'expression trouvée (27) pour le coefficient de corrélation père-fils est, en fonction de la variance totale V et de la variance génétique V_0 , la même que celle qu'a établie M. MALÉCOT dans le cas du random mating ⁽¹⁾. Mais les expressions de ces variances en fonction des variances partielles $M(H^2)$ et $M(G^2)$ ne sont pas les mêmes.

Or, nous avons vu que les valeurs des aléatoires H et G sont identiques (avec l'approximation convenue) à celles que l'on aurait dans le cas du random mating : il est donc possible de comparer dans une population en équilibre, le coefficient de corrélation père-fils à celui que l'on obtiendrait dans la même population soumise au random mating.

Désignons par δ_0^2 le rapport de dominance (« dominance ratio ») :

$$\delta_0^2 = \frac{\Sigma M(G^2)}{\Sigma M(H^2)}$$

si l'on désigne par ρ_0 le coefficient de corrélation père-fils dans le cas du random mating, on a :

$$\rho_0 = \delta_0^2/2$$

Posons d'autre part :

$$\tau^2 = \Sigma M(G^2); \quad \sigma^2 = \Sigma M(H^2); \quad \text{d'où :} \quad \delta_0^2 = \tau^2/\sigma^2$$

Les expressions (15') et (20') deviennent :

$$\begin{cases} V_0 = \tau^2 - 2\mu V_0^2 \\ V = \sigma^2 - 2\mu V_0^2 \end{cases}$$

(1) Cf. G. MALÉCOT, *Thèse*, Paris, 1939, p. 51.

Et le coefficient de corrélation ρ s'écrit :

$$(27') \quad \rho = 1/2 [\tau^2/\sigma^2 + 2 \mu V_0^2 (\tau^2 - \sigma^2)/\sigma^4]$$

Il apparaît que ρ est toujours inférieur ou égal à ρ_0 , l'égalité correspondant au cas où il n'y a pas de dominance.

Donc, un effet de la loi d'union sélective adoptée est de réduire le coefficient de corrélation père-fils s'il y a dominance.

Précisons davantage. L'équation (27') s'écrit en tenant compte de $V_0 = \tau^2 + 0 (\mu)$:

$$\rho = \rho_0 + \mu (\tau^2 - \sigma^2) \tau^2/\sigma^4$$

et, comme :

$$\tau^2 - \sigma^2 = (2 \rho_0 - 1) \sigma_0$$

On obtient :

$$\rho = \rho_0 [1 + 4 \mu \rho_0 (2 \rho_0 - 1) \sigma^2]$$

et finalement :

$$(28) \quad \rho/\rho_0 = 8 \mu \rho_0^2 \sigma^2 - 4 \mu \rho_0 \sigma^2 + 1$$

Le rapport ρ/ρ_0 qui est maximum et égal à l'unité lorsque $\rho_0 = 1/2$ (absence de dominance) présente (pour une valeur donnée de σ^2) un minimum lorsque

$$\rho_0 = 1/4$$

En résumé donc, pour une population donnée en équilibre, le coefficient de corrélation père-fils diffère, s'il y a dominance, de celui que l'on aurait dans le cas du random mating : il est diminué ; et cette diminution est plus sensible lorsque le coefficient est voisin de $1/4$: on a alors sensiblement :

$$(29) \quad \rho/\rho_0 = 1 - \mu \sigma^2/2$$

*
* *

Corrélation entre frères :

Nous désignerons par des lettres affectées des indices 1 et 2 les aléatoires relatives aux deux frères, les lettres sans indice représentant les aléatoires analogues relatives aux parents.

M. MALÉCOT a montré ⁽¹⁾ que, si l'on fixe les constitutions génétiques des parents, on a :

$$M(H_1 H'_2) = M(H_1 H'_1) \quad \text{et} \quad M(H_1 H_2) = [M(H_1)]^2$$

(1) Cf. G. MALÉCOT, *Thèse*, Paris, 1939, p. 92.

Il en résulte que la table d'association des phases du facteur (H) chez l'un des frères et du facteur (H') chez l'autre frère est la même que la table d'association des phases de ces deux facteurs chez un même individu. D'autre part, pour calculer la moyenne $M(H_1H_2)$, il suffit de classer les parents suivant leurs génotypes et d'écrire les fréquences des enfants de génotype donné issus des différentes catégories de couples : ces fréquences sont données dans le tableau III suivant :

TABLEAU III

Enfants Parents		AA	Aa	aa
père	AA	$p^2 * p^2$	0	0
mère	AA			
père	AA	$1/2 p^2 * 2 pq$	$1/2 p^2 * 2 pq$	0
mère	Aa			
père	AA	0	$p^2 * q^2$	0
mère	aa			
père	Aa	$1/2 2 pq * p^2$	$1/2 2 pq * p^2$	0
mère	AA			
père	Aa	$1/4 2 pq * 2 pq$	$1/2 2 pq * 2 pq$	$1/4 2 pq * 2 pq$
mère	Aa			
père	Aa	0	$1/2 2 pq * q^2$	$1/2 2 pq * q^2$
mère	aa			
père	aa	0	$q^2 * p^2$	0
mère	AA			
père	aa	0	$1/2 q^2 * 2 pq$	$1/2 q^2 * 2 pq$
mère	Aa			
père	aa	0	0	$q^2 * q^2$
mère	aa			

Or, on a :

$$M(H_1x_2) = M(H_1H_2) + \Sigma' M(H_1H'_2)$$

et, d'après la remarque que nous avons faite :

$$M(H_1x_2) = M(H_1H_2) + \Sigma' M(H_1H'_1)$$

ou encore :

$$M(H_1x_2) = M(H_1H_2) + M(H_1x_1) - M(H_1^2)$$

Et, en posant :

(30)

$$g = M(H_1^2) - M(H_1H_2)$$

on obtient :

(31)

$$M(x_1 x_2) = \Sigma M(H_1 x_2) = M(x_1^2) - \Sigma g$$

Calcul de g :

Le tableau III permet, en raison de la remarque faite au début de ce paragraphe, d'écrire l'expression de la moyenne $M(H_1 H_2)$:

$$\begin{aligned} M(H_1 H_2) = & (p^2 * p^2) i^2 + (p^2 * 2pq)(i+j)^2/4 + (p^2 * q^2) j^2 \\ & + (2pq * p^2)(i+j)^2/4 + (2pq * 2pq)(i+2j+k)^2/16 \\ & + (2pq * q^2)(j+k)^2/4 + (q^2 * p^2) j^2 + (q^2 * 2pq)(j+k)^2/4 \\ & + (q^2 * q^2) k^2 \end{aligned}$$

On a de même l'expression de $M(H_1^2)$:

$$\begin{aligned} M(H_1^2) = & (p^2 * p^2) i^2 + (p^2 * 2pq)(i^2 + j^2)/2 + (p^2 * q^2) j^2 \\ & + (2pq * p^2)(i^2 + j^2)/2 + (2pq * 2pq)(i^2 + 2j^2 + k^2)/4 \\ & + (2pq * q^2)(j^2 + k^2)/2 + (q^2 * p^2) j^2 + (q^2 * 2pq)(j^2 + k^2)/2 \\ & + (q^2 * q^2) k^2 \end{aligned}$$

d'où l'on déduit :

$$g = (pq * p)(i-j)^2 + (pq * q)(j-k)^2 - (pq * pq)(i-2j+k)^2/4$$

et en remarquant que :

$$pq * p = p^2 q (1 + \mu y_{200}) + 0(\mu^2) \quad \text{car} \quad y_{100} = T = 0(\mu)$$

$$pq * q = pq^2 (1 + \mu y_{200}) + 0(\mu^2) \quad \text{car} \quad y_{300} = S = 0(\mu)$$

et :

$$pq * pq = p^2 q^2 (1 + 2\mu y_{200})$$

on obtient :

$$\begin{aligned} g = & pq \left\{ p(i-j)^2 + q(j-k)^2 - \frac{pq(i-2j+k)^2}{4} \right. \\ & \left. + \mu y_{200} [p(i-j)^2 + q(j-k)^2 - \frac{pq(i-2j+k)^2}{2}] \right\} + 0(\mu^2) \end{aligned}$$

Introduisons alors les résidus de dominance m, n, r , et posons :

$$i_0 - j_0 = j_0 - k_0 = e$$

d'où :

$$\begin{cases} i - j = e + m - n \\ j - k = e + n - r \end{cases}$$

Et comme la condition d'ajustement (1) donne :

$$pm + qn = 0 (\mu^2); \quad pn + qr = 0 (\mu^2)$$

on peut poser, avec l'approximation adoptée :

$$(32) \quad p^2m = -pqn = q^2r = p^2q^2C$$

d'où :

$$(32') \quad \begin{array}{lll} m = q^2C; & n = -pqC; & r = p^2C \\ m - n = qC; & n - r = -pC \end{array}$$

et en écrivant :

$$i - j = e + qC; \quad j - k = e - pC; \quad i - 2j + k = C$$

l'expression trouvée pour g devient :

$$(33) \quad g = pq \left[e^2 + \frac{3pqC^2}{4} + \mu y_{200} \left(e^2 + \frac{pqC^2}{2} \right) \right]$$

Or e et C sont de l'ordre de grandeur des contributions de chaque facteur ; si le nombre n de ceux-ci est assez grand, le coefficient de μ dans (33) est de l'ordre de $1/n^2$. On a donc dans ce cas :

$$g = pq \left(e^2 + \frac{3pqC^2}{4} \right) + 0 (\mu/n^2)$$

et on obtient pour Σg l'expression approchée :

$$(34) \quad \Sigma g = \Sigma pq \left(e^2 + \frac{3pqC^2}{4} \right)$$

Pour interpréter cette expression, remarquons avec M. MALÉCOT, que les contributions génétiques sont proportionnelles à e (1) ; en effet, les relations :

$$pio + qk_o = 0 \quad [\text{conséquence de (2)}]$$

$$\text{et } i_o - k_o = 2e$$

entraînent

$$(35) \quad i_o = 2qe; \quad j_o = (q - p)e; \quad k_o = -2pe$$

(1) Cf. G. MALÉCOT, *Thèse*, Paris, 1939, p. 94.

On a donc :

$$M(G^2) = p^2 i_0^2 + 2 p q j_0^2 + q^2 k_0^2 = 2 p q e^2$$

On a de même d'après (32') :

$$M(d^2) = p^2 m^2 + 2 p q n^2 + q^2 r^2 = p^2 q^2 C^2$$

et l'équation (34) devient :

$$(34') \quad \Sigma g = \Sigma M(G^2)/2 + 3 \Sigma M(d^2)/4$$

L'équation (31) donne alors :

$$M(x_1 x_2) = M(x_1^2) - \Sigma M(G^2)/2 - 3 \Sigma M(d^2)/4$$

ce qui s'écrit finalement d'après (15') et (21) :

$$(36) \quad M(x_1 x_2) = V - V_0/2 - \mu V_0^2 - 3 M(D^2)/4 + 0 (\mu^2)$$

On en déduit, en tenant compte de $M(D^2) = V - V_0$, la valeur du coefficient de corrélation entre frères :

$$\varphi = M(x_1 x_2)/V = 1 - V_0/2V - \mu V_0^2/V - 3(V - V_0)/4V$$

ou, en posant :

$$\delta^2 = V_0/V$$

$$(37) \quad \varphi = \frac{1}{4} + \frac{\delta^2}{4} - \mu \delta^2 V_0$$

Remarque. — L'expression (37) peut encore s'écrire en posant :

$$(37') \quad \begin{aligned} A &= -2 \mu V_0 \\ \varphi &= \delta^2 (1 + A)/2 + (1 - \delta^2)/4 \end{aligned}$$

Il est curieux de constater que cette expression est identique à celle qui a été établie par M. MALÉCOT et prévue par M. Fisher dans le cas d'homogamie qu'ils ont étudié (1). Mais la quantité A ne représente plus ici le coefficient de corrélation entre z et $\bar{z}$, puisque nous avons vu que ce dernier est nul.

(1) Cf. G. MALÉCOT, *Thèse*, Paris, 1939, p. 95.

Comparaison avec le random mating :

Contrairement à ce que nous avons trouvé pour la corrélation père-fils, le coefficient φ n'a pas la même expression en fonction des variances V et V_0 que dans le cas du random mating. On trouve en effet dans ce cas :

$$\varphi_0 = (1 + \delta_0^2)/4 = \delta_0^2/2 + (1 - \delta_0^2)/4$$

En particulier, si nous considérons le cas où il n'y a pas de dominance, nous avons vu que les coefficients de corrélation père-fils ρ et ρ_0 sont les mêmes ; les coefficients de corrélation entre frères φ et φ_0 sont au contraire différents ; leurs valeurs sont :

$$(38) \quad \begin{array}{l} \varphi_0 = 1/2 \\ \varphi = (1 - 2\mu V)/2 \end{array}$$

Le mode d'union sélective adopté entraîne donc une diminution du coefficient de corrélation entre frères.

Remarque. — Il est possible d'établir directement l'expression (38) du coefficient de corrélation entre frères dans le cas où il n'y a pas de dominance.

En effet, si on fixe la constitution génétique des parents, les constitutions des enfants deviennent stochastiquement indépendantes et on a :

$$M(x_1x_2) = [M(x_1)]^2 = (x + \bar{x})^2/4$$

Si l'on désigne par π et $\bar{\pi}$ les fréquences des catégories dans lesquelles se trouvent les parents et si l'on envisage tous les cas possibles, on a :

$$M(x_1x_2) = \Sigma \Sigma \pi \bar{\pi} (1 + \mu y + \mu \bar{y}) (x + \bar{x})^2/4$$

d'où :

$$\begin{aligned} M(x_1x_2) = & M(x^2)/4 + M(\bar{x}^2)/4 + M(x\bar{x})/2 + \mu M(x^2y)/4 + \mu M(x^2\bar{y})/4 \\ & + \mu M(x\bar{x}y)/2 + \mu M(x\bar{x}\bar{y})/2 + \mu M(\bar{x}^2y)/4 + \mu M(\bar{x}^2\bar{y})/4 \end{aligned}$$

Or, puisque il n'y a pas de corrélation entre conjoints :

$$M(x\bar{x}) = M(x^2\bar{y}) = M(x\bar{x}y) = M(x\bar{x}\bar{y}) = M(\bar{x}^2y) = 0$$

et il vient finalement :

$$M(x_1x_2) = M(x^2)/2 + \mu M(x^2y)/2$$

soit en remplaçant y par son expression (IV-11) en fonction de x :

$$\begin{aligned} y = & -x^2 + 2 \underline{O}x + V \\ M(x_1x_2) = & V/2 - [M(x^4) - 2 \underline{O}M(x^3) - V^2] \mu/2 \end{aligned}$$

Les contributions des différents facteurs étant, selon notre hypothèse, petites et les facteurs très nombreux, l'effectif de la population n'étant d'autre part pas limité, on peut supposer que la loi de répartition de x est sensiblement gaussienne, et on a alors :

$$M(x^4) = 3 V^2; \quad M(x^3) = 0$$

d'où :

$$M(x_1 x_2) = V(1 - 2\mu V)/2$$

et l'on retrouve bien l'expression (38) de φ .

CHAPITRE VI

FECONDITÉ ET SURVIVANCE DIFFÉRENTIELLES UNION PANMICTIQUE APRÈS SÉLECTION

Fécondité différentielle :

Considérons une population soumise au « random mating », mais dont les couples n'ont pas tous la même fécondité. La fréquence des couples formés de deux individus pris respectivement dans une catégorie de fréquence π et une catégorie de fréquence $\bar{\pi}$ est :

$$\pi \cdot \bar{\pi}$$

Supposons alors de façon plus précise que la fréquence des enfants issus de tels couples dans la génération suivante soit :

$$(1) \quad \pi \cdot \bar{\pi} (1 + \mu y + \mu \bar{y})$$

y et $\bar{y}$ représentant la valeur centrée d'un caractère H dans les catégories considérées.

On voit que la répartition des enfants, et donc des gènes et des génotypes dans chaque génération, est la même que si la population était soumise à la loi d'union sélective que nous avons considérée jusqu'alors, les couples ayant tous la même fécondité.

*
* *
*

Survivance différentielle :

Considérons une population panmictique ; mais supposons que les enfants n'ont pas tous la même probabilité d'atteindre l'âge adulte. De façon plus précise, supposons que, si l'on désigne par π_0 la fréquence des individus d'un type donné dans la génération *naissante*, ces individus aient dans la génération *adulte* la fréquence :

$$(2) \quad \pi = \pi_0 (1 + \mu y_0)$$

y_0 étant la valeur centrée d'un caractère H pour le type d'individus considéré, dans la génération *naissante*.

Remarquons que l'on a :

$$\Sigma \pi = \Sigma \pi_0 = 1$$

D'autre part, si au lieu de considérer des individus identiques, on envisage un ensemble d'individus de fréquence totale :

$$\pi' = \Sigma' \pi$$

on a :

$$\pi' = \Sigma' \pi = \Sigma' \pi_0 (1 + \mu y_0) = \Sigma' \pi_0 + \mu \Sigma' \pi_0 y_0 = \pi_0' (1 + \mu y_0')$$

en désignant par y_0' la valeur moyenne centrée du caractère dans la catégorie naissante considérée.

Puisqu'on a supposé les croisements panmictiques, la probabilité pour que les conjoints d'un couple proviennent respectivement de deux catégories déterminées de fréquences π et $\bar{\pi}$ est :

$$\pi \cdot \bar{\pi} = \pi_0 \cdot \bar{\pi}_0 (1 + \mu y_0) (1 + \mu \bar{y}_0) = \pi_0 \cdot \bar{\pi}_0 (1 + \mu y_0 + \mu \bar{y}_0) + 0 (\mu^2)$$

Donc, avec l'approximation adoptée, tout se passe au point de vue des générations *naissantes* comme si les enfants avaient tous même probabilité de survivance, la population étant soumise à la loi d'union sélective que nous avons envisagée jusqu'ici.

*
* *

Population panmictique soumise simultanément à une fécondité et une survivance différentielle :

Supposons une population panmictique soumise d'une part à une loi de fécondité différentielle du type (1) de coefficient μ' , et d'autre part à une loi de survivance différentielle du type (2) de coefficient μ'' .

Avec les notations précédentes, la fréquence dans la génération naissante, des enfants issus de couples dont les conjoints sont respectivement des individus dont le type a pour fréquence π et $\bar{\pi}$, est :

$$\begin{aligned} \pi \cdot \bar{\pi} (1 + \mu' y + \mu' \bar{y}) &= \pi_0 \cdot \bar{\pi}_0 (1 + \mu'' y_0 + \mu'' \bar{y}_0) (1 + \mu' y + \mu' \bar{y}) \\ &= \pi_0 \cdot \bar{\pi}_0 (1 + \mu'' y_0 + \mu' y + \mu'' \bar{y}_0 + \mu' \bar{y}) + 0 (\mu^2) \end{aligned}$$

en désignant par $0 (\mu^2)$ l'ensemble des termes négligeables $0 (\mu'^2)$, $0 (\mu''^2)$, $0 (\mu' \mu'')$.

Or, si H est la valeur du caractère pour un individu, les valeurs centrées de ce caractère pour le même individu dans la génération naissante et dans la génération adulte sont :

$$y_0 = H - \Sigma \pi_0 H \quad \text{et} \quad y = H - \Sigma \pi H$$

d'où :

$$y - y_0 = \Sigma \pi_0 H - \Sigma \pi H = \Sigma \pi_0 H - \Sigma \pi_0 (1 + \mu'' y_0) H = 0 (\mu'')$$

Donc la fréquence des enfants issus des couples considérés est, avec l'approximation adoptée :

$$\pi_0 \cdot \bar{\pi}_0 [1 + (\mu' + \mu'') y_0 + (\mu' + \mu'') \bar{y}_0]$$

Pour la génération naissante, tout se passe donc comme si la population ne présentait ni fécondité différentielle, ni survivance différentielle, mais était soumise à la loi d'union sélective considérée jusqu'alors, de coefficient :

$$\mu = \mu' + \mu''$$

Remarque. — Les résultats que nous avons établis pour l'évolution d'une population soumise à une loi d'union sélective sont encore valables dans les hypothèses d'une population panmictique soumise à une loi de fécondité différentielle ou à une loi de survivance différentielle des types indiqués, à condition de considérer les générations naissantes.

En ce qui concerne les corrélations entre apparentés, seule la corrélation entre frères dans une génération naissante d'une population en équilibre a encore la forme établie. Les corrélations père-fils ne sont évidemment plus les mêmes puisque les populations adultes n'ont pas la même composition.

*
* * *

Comparaison avec l'hypothèse de M. Sewall Wright (1) :

Proposons-nous de comparer la loi de survivance différentielle (ou encore d'union panmictique après sélection) définie par (2) avec la loi de sélection considérée par M. S. WRIGHT.

Rappelons tout d'abord pour cela, avec des notations analogues à celles qu'il a utilisées, les hypothèses faites par M. S. WRIGHT.

Le caractère primaire S, supposé sans dominance, est la somme des contributions α_i des gènes A_i ; sa moyenne est :

$$M(S) = \bar{M} = 2 \sum \alpha_i p_i$$

Le caractère secondaire H (niveau d'adaptation) relatif à S est la somme d'une partie génétique, somme des contributions génétiques γ_i des différents gènes A_i (les contributions des gènes a_i pouvant être considérées comme nulles) et des résidus de dominance δ_i correspondant à chaque facteur, ainsi que des résidus d'épistasy ϵ_{ij} correspondant à chaque couple de facteurs. On a donc :

$$H = G + D + E$$

en posant :

$$G = \sum \gamma_i ; \quad D = \sum \delta_i ; \quad E = \sum \epsilon_{ij}$$

La valeur moyenne du caractère secondaire H est :

(3)

$$M(H) = \bar{H} = 2 \sum \alpha_i^2 p_i q_i + (M - \bar{O})^2$$

(1) Sewall WRIGHT, « The analysis of variance and correlations between relatives with respect to deviations from an optimum », *J. of Genetics*, vol. XXX, 1935, p. 243-256.

Id., « Evolution in populations in approximate equilibrium », *J. of Genetics*, vol. XXX, 1935, p. 257-266.

Les résidus de dominance δ_i sont calculés au moyen d'un ajustement par moindres carrés, chaque quantité $\sigma^2 = M(H - G)^2$ relative à un seul facteur étant rendue minimum. Les valeurs trouvées vérifient :

$$M(D) = 0$$

En remarquant que l'on a :

$$\frac{\partial M(G)}{\partial p_i} = 2 \gamma_i$$

M. S. WRIGHT calcule les contributions génétiques γ_i , en supposant que l'on a :

$$(4) \quad M(E) = 0$$

ce qui entraîne :

$$M(H) = M(G)$$

et donc :

$$\frac{\partial \bar{H}}{\partial p_i} = 2 \gamma_i$$

L'expression (3) donne alors :

$$\gamma_i = -\alpha_i [\alpha_i (1 - 2 p_i) + 2 (M - \underline{O})]$$

La variation du niveau d'adaptation $\bar{H}$ de l'ensemble de la population est alors :

$$\Delta \bar{H} = \sum \frac{\partial \bar{H}}{\partial p_i} \Delta p_i = \sum 2 \gamma_i \Delta p_i$$

Remarquons d'ailleurs que l'hypothèse (4) est justifiée ; en effet, comme l'indique M. S. WRIGHT, le calcul direct des résidus d'epistacy par des relations telles que :

$$\epsilon_{A_1 A_1 A_2 A_2} = (\bar{H}_{A_1 A_1 A_2 A_2} - \bar{H}) - (\bar{H}_{A_1 A_1} - \bar{H}) - (\bar{H}_{A_2 A_2} - \bar{H})$$

permet de vérifier (4).

Ces résultats établis, M. S. WRIGHT écrit : « In the general case the selective disadvantage must be proportional to the momentary net effect of gene replacement on adaptive value », et pose ⁽¹⁾ :

$$(5) \quad \Delta p_i = s \gamma_i p_i q_i$$

ce qui conduit à :

$$\Delta \bar{H} = 2 s \sum \gamma_i^2 p_i q_i$$

(1) Cf. S. WRIGHT, *op. cit.*, p. 258, équation (4).

Reprenons alors notre hypothèse (2) de loi d'union panmictique après sélection.

La loi de Hardy étant vérifiée pour les générations naissantes, les valeurs de δ_i établies par M. S. WRIGHT sont encore valables et on a encore $M(D) = 0$.

On a de même l'expression (3).

Toutefois, par suite de la liaison introduite entre les facteurs par notre loi de sélection (on a par exemple $(II) = p_1^2 p_2^2 + 0(\mu)$ et non plus $(II) = p_1^2 p_2^2$), les valeurs calculées par M. WRIGHT pour les résidus d'épistasy ne vérifient pas la condition (4) et on a :

$$M(E) = 0(\mu)$$

par conséquent :

$$\frac{\partial \bar{H}}{\partial p_i} = \frac{\partial M(G)}{\partial p_i} + 0(\mu)$$

et donc :

$$\gamma_i = -\alpha_i [\alpha_i (1 - 2p_i) + 2(M - \bar{O})] + 0(\mu)$$

On déduit alors de (3) :

$$\Delta \bar{H} = \sum \frac{\partial \bar{H}}{\partial p_i} \Delta p_i = \sum 2 \gamma_i \Delta p_i + \sum 0(\mu) \Delta p_i$$

ou encore, avec l'approximation adoptée, comme $\Delta p = 0(\mu)$, on a :

$$\Delta \bar{H} = \sum 2 \gamma_i \Delta p_i + 0(\mu^2)$$

On obtient, pour l'expression de la variation de niveau d'adaptation, la même expression que M. WRIGHT. Pour comparer les hypothèses faites, il faut donc comparer les expressions de Δp_i .

Désignons par P_i , $2 Q_i$, R_i les fréquences des génotypes $A_i A_i$, $A_i a_i$, $a_i a_i$ dans la génération adulte, par p_i et q_i les fréquences des gènes dans la génération adulte précédente, c'est-à-dire aussi, puisqu'il y a random mating, dans la même génération naissante. On a d'après (2) :

(6)

$$\begin{aligned} P_i &= p_i^2 (1 + \mu y_{A_i A_i}) \\ 2 Q_i &= 2 p_i q_i (1 + \mu y_{A_i a_i}) \\ R_i &= q_i^2 (1 + \mu y_{a_i a_i}) \end{aligned}$$

où $y_{A_i A_i}$ est la valeur moyenne centrée de H pour la catégorie de génotype $A_i A_i$ dans la génération naissante, etc...

On a :

$$\Delta p_i = P_i + Q_i - p_i$$

Or (6) s'écrit encore :

$$(7) \quad \begin{cases} P_i/p_i^2 - Q_i/p_i q_i = \mu (y_{A_i A_i} - y_{A_i a_i}) \\ Q_i/p_i q_i - R_i/q_i^2 = \mu (y_{A_i a_i} - y_{a_i a_i}) \\ P_i + 2 Q_i + R_i = 1 \end{cases}$$

Calculons les différences figurant aux deuxièmes membres des deux premières équations (7) :

$$\begin{aligned} y_{A_i A_i} - y_{A_i a_i} &= H_{A_i A_i} - H_{A_i a_i} \\ &= G_{A_i A_i} - G_{A_i a_i} + \delta_{A_i A_i} - \delta_{A_i a_i} + \epsilon_{A_i A_i} - \epsilon_{A_i a_i} \end{aligned}$$

en désignant par $\epsilon_{A_i A_i}, \dots$ les valeurs moyennes des résidus d'épistasy correspondant à tous les couples formés par un facteur autre que i et le facteur i sous sa phase $A_i A_i$, etc...

Ces résidus d'épistasy étant ceux qui ont été calculés par M. S. WRIGHT, on a par exemple :

$$\epsilon_{A_i A_i} = \sum \frac{p_i^2 p_j^2 (-8 \alpha_i \alpha_j q_i q_j) + 2 p_i^2 p_j q_j [-4 \alpha_i \alpha_j q_i (1 - 2 p_i)] + p_i^2 q_j^2 (8 \alpha_i \alpha_j p_i q_i)}{p_i^2} + 0 (\mu)$$

la somme étant étendue à toutes les valeurs de j autres que i . Le calcul montre immédiatement que cette expression se réduit à 0 (μ).

D'autre part, les résultats établis par M. WRIGHT donnent :

$$G_{A_i A_i} - G_{A_i a_i} = \gamma_i; \quad \delta_{A_i A_i} - \delta_{A_i a_i} = -2 \alpha_i^2 q_i$$

et de même :

$$G_{A_i a_i} - G_{a_i a_i} = \gamma_i; \quad \delta_{A_i a_i} - \delta_{a_i a_i} = 2 \alpha_i^2 p_i$$

d'où finalement :

$$\begin{cases} y_{A_i A_i} - y_{A_i a_i} = \gamma_i - 2 \alpha_i^2 q_i + 0 (\mu) \\ y_{A_i a_i} - y_{a_i a_i} = \gamma_i + 2 \alpha_i^2 p_i + 0 (\mu) \end{cases}$$

Le système (7) donne alors :

$$\begin{cases} P_i = p_i^2 [1 + 2 \mu q_i (\gamma_i - \alpha_i^2 q_i)] + 0 (\mu^2) \\ Q_i = p_i q_i \left\{ 1 + \mu [2 \alpha_i^2 p_i q_i - \gamma_i (p_i - q_i)] \right\} + 0 (\mu^2) \\ R_i = q_i^2 [1 - 2 \mu p_i (p_i \alpha_i^2 + \gamma_i)] + 0 (\mu^2) \end{cases}$$

d'où :

$$p_i + \Delta p_i = P_i + Q_i = p_i (1 + \mu \gamma_i q_i) + 0 (\mu^2)$$

et finalement :

$$(8) \quad \Delta p_i = \mu \gamma_i p_i q_i + 0 (\mu^2)$$

L'expression (8) est précisément, avec notre approximation, celle que M. S. WRIGHT a prise pour hypothèse (5).

En résumé donc, l'hypothèse de sélection adoptée par M. S. WRIGHT, conduit, avec l'approximation que nous avons admise, au point de vue de l'évolution, aux mêmes résultats que notre loi d'union sélective.

Remarque. — Dans le cas où le caractère primaire ne dépend que d'un seul facteur, nous avons vu (II-2) que :

$$\Delta p = \mu p T = \mu p q [2 \alpha \underline{O} - \alpha^2 (1 + 2 p)]$$

Or on a immédiatement :

$$p q \gamma = p q [-\alpha^2 (1 - 2 p) - 2 \alpha (2 \alpha p - \underline{O})] = p q [2 \alpha \underline{O} - \alpha^2 (1 + 2 p)]$$

On vérifie bien :

$$p T = \gamma p q$$

DEUXIÈME PARTIE

POPULATION D'EFFECTIF LIMITÉ CONSTANT

CHAPITRE PREMIER

LOI DE PROBABILITÉ DES FRÉQUENCES DES GÈNES AU COURS DES GÉNÉRATIONS

EXISTENCE ET FORME DE LA LOI ASYMPTOTIQUE

Dans cette deuxième partie de notre étude, nous supposerons que la population soumise à la loi d'union sélective précédemment définie est limitée et que son effectif reste constant au cours des générations. Nous désignerons par N le nombre d'individus de chaque génération. Enfin, la population considérée sera monoïque, l'autofécondation étant possible mais ni plus ni moins probable que la fécondation croisée.

Pour ne pas alourdir les notations, nous désignerons simplement par des lettres non accentuées ce qui concerne la n^{me} génération F_n et par des lettres accentuées ce qui se rapporte à la génération suivante F_{n+1} .

Du fait de la limitation de l'effectif, les fréquences $p, q, P, 2Q, R$ des gènes et des génotypes sont des aléatoires, en chaînes simples de Markoff. Remarquons tout d'abord que, contrairement à ce qui a lieu dans le cas du random mating, la loi de probabilité de la fréquence p' ne dépend pas seulement de la fréquence p , mais aussi des fréquences des génotypes $P, 2Q, R$. C'est ainsi par exemple que la valeur moyenne de p' est :

$$p + \mu p T$$

la quantité T dépendant des fréquences des génotypes dans la génération F_n . Il sera donc nécessaire pour étudier la loi de probabilité de p , d'étudier les lois de probabilités simultanées des différentes variables $p, q, P, 2Q, R$.

D'autre part, la connaissance d'une des fréquences p, q et d'une des fréquences $P, 2Q, R$ entraîne la connaissance des autres et par conséquent celle de la génération F_n . Nous étudierons donc la loi de probabilité simultanée des deux aléatoires (non indépendantes) p et P .

Pour cela nous nous proposons de déterminer la fonction caractéristique $F(x, y, t)$ de cette loi à la date t ; nous avons par définition :

$$(1) \quad F(x, y, t) = M(e^{xp + yP}) \quad (1)$$

Si nous désignons par $\Delta F, \Delta p, \Delta P$ les variations de F, p, P correspondant à l'intervalle de temps Δt , on a évidemment :

$$\begin{aligned} F + \Delta F &= M[e^{x(p + \Delta p) + y(P + \Delta P)}] = M(e^{xp + yP} \cdot e^{x\Delta p + y\Delta P}) \\ &= M[e^{xp + yP} \cdot M_c(e^{x\Delta p + y\Delta P})] \end{aligned}$$

M_c représentant la moyenne conditionnée, c'est-à-dire lorsque les fréquences p et P ont une valeur déterminée ; et finalement :

$$(2) \quad F + \Delta F = M \left\{ e^{xp + yP} \cdot M_c [1 + (x\Delta p + y\Delta P) + (x\Delta p + y\Delta P)^2/2! + \dots] \right\}$$

Posons alors :

$$\begin{aligned} M_c(\Delta p) &= A_{10}(p, P, t) \Delta t ; & M_c(\Delta P) &= A_{01}(p, P, t) \Delta t \\ M_c[(\Delta p)^2] &= B_{20}^2(p, P, t) \Delta t ; & M_c[(\Delta P)^2] &= B_{02}^2(p, P, t) \Delta t ; \\ M_c(\Delta p \cdot \Delta P) &= B_{11}^2(p, P, t) \Delta t \\ M_c[(\Delta p)^3] &= C_{30}^3(p, P, t) \Delta t ; & M_c[(\Delta p)^2 \cdot \Delta P] &= C_{21}^3(p, P, t) \Delta t ; \text{ etc...} \end{aligned}$$

L'équation (2) donne :

$$\Delta F = M \left\{ e^{xp + yP} \cdot M_c [(x\Delta p + y\Delta P) + (x\Delta p + y\Delta P)^2/2! + \dots] \right\}$$

d'où :

$$(3) \quad \Delta F = \Delta t \cdot M \left\{ e^{xp + yP} [A_{10}x + A_{01}y + (B_{20}^2x^2 + 2B_{11}^2xy + B_{02}^2y^2)/2 + \dots] \right\}$$

Convenons alors de poser de façon symbolique :

$$M \left[f(p, P) \cdot e^{xp + yP} \right] = f\left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right) F \quad (2)$$

(1) P et p variant entre 0 et 1, il ne se pose pas de question de convergence, et il est inutile de prendre l'exposant i ($xp + yP$).

(2) Cf. M. S. BARTLETT, « Some Evolutionary Stochastic Processes », *Journal of the Royal Statistical Society*, vol. XI, n° 2, 1949, p. 220.

Et G. MALÉCOT, *Les processus stochastiques et la méthode des fonctions génératrices ou caractéristiques*, p. 7 (Publications de l'Institut de Statistique, Paris, vol. I, fasc. III, 1952).

c'est-à-dire que l'on a en particulier :

$$(4) \quad \frac{\partial F}{\partial x} = M(p \cdot e^{px} + Py); \quad \frac{\partial F}{\partial y} = M(P \cdot e^{py} + Px)$$

$$\frac{\partial^2 F}{\partial x^2} = M(p^2 \cdot e^{px} + Py); \quad \frac{\partial^2 F}{\partial x \partial y} = M(pP \cdot e^{px} + Py); \quad \frac{\partial^2 F}{\partial y^2} = M(P^2 \cdot e^{py} + Px)$$

L'équation (3) s'écrit alors :

$$(5) \quad \frac{\partial F}{\partial t} = xA_{10} \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, t \right) F + yA_{01} \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, t \right) F + x^2/2 B_{20} \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, t \right) F$$

$$+ xyB_{11} \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, t \right) F + y^2/2 B_{02} \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, t \right) F$$

$$+ \dots$$

Le problème revient donc à calculer les moyennes conditionnées A, B²...

* * *

Remarquons que si l'on prend comme intervalle de temps Δt une génération, on a :

$$M_c(\Delta P) = M_c(P' - P) = M_c(P') - P$$

et de même :

$$M_c(\Delta p) = M_c(p') - p$$

Les résultats établis dans la première partie (pages 5 et 6) permettent d'écrire :

$$M_c(p') = p(1 + \mu T) \quad \text{et} \quad M_c(P') = p * p = p^2(1 + 2\mu T)$$

Donc :

$$M_c(\Delta p) = \mu p T; \quad M_c(\Delta P) = p^2(1 + 2\mu T) - P$$

Si l'on convient de prendre comme unité de temps la génération, on a donc :

$$A_{10} = \mu p T \quad \text{et} \quad A_{01} = p^2(1 + 2\mu T) - P$$

ou, plus généralement :

$$(6) \quad A_{10}\Delta t = \mu p T \Delta t \quad A_{01}\Delta t = [p^2(1 + 2\mu T) - P] \Delta t$$

Si nous considérons maintenant les variances conditionnées, nous pouvons évidemment écrire :

$$\sigma_c^2(\Delta p) = \sigma_c^2(p') \quad \text{et} \quad \sigma_c^2(\Delta P) = \sigma_c^2(P')$$

si l'on prend pour intervalle de temps Δt une génération, car on a alors :

$$\Delta p = p' - p \quad \text{et} \quad \Delta P = P' - P, p \text{ et } P \text{ étant déterminés.}$$

Or, si l'on suppose que les N individus qui composent la génération F_{n+1} sont choisis au hasard parmi tous les enfants qui auraient pu être issus de la génération F_n , c'est-à-dire qu'il n'y a pas d'autre sélection que celle due à l'union sélective, le nombre des individus de génotype AA dans F_{n+1} a une variance donnée par la formule de Bernouilli :

$$M_c(P') = [1 - M_c(P')] \cdot N$$

La variance conditionnée de la fréquence P' est donc :

$$\sigma_c^2(P') = M_c(P') [1 - M_c(P')] / N$$

Cette variance est égale à la variance conditionnée de ΔP correspondant à $\Delta t = 1$. Plus généralement, on a :

$$\sigma_c^2(\Delta P) = M_c(P') [1 - M_c(P')] \Delta t / N$$

ou :

$$\sigma_c^2(\Delta P) = p^2 (1 + 2 \mu T) [1 - p^2 (1 + 2 \mu T)] \Delta t / N$$

Or :

$$\sigma_c^2(\Delta P) = M_c [(\Delta P)^2] - [M_c(\Delta P)]^2$$

ou :

$$M_c [(\Delta P)^2] = \sigma_c^2(\Delta P) + A_{01}^2 (\Delta t)^2$$

Si l'on considère comme négligeables les termes $0 (\Delta t)^2$ on a :

$$B_{02}^2 \Delta t = \sigma_c^2(\Delta P)$$

et si N est suffisamment grand pour que $0 (\mu/N)$ soit négligeable comme $0 (\mu^2)$ on trouve finalement :

(7)

$$B_{02}^2 \Delta t = p^2 (1 - p^2) \Delta t / N$$

On a de même pour $\Delta t = 1$:

$$M_c(\Delta Q) = M_c(Q') - Q = p * (1 - p) - Q = p (1 - p) (1 + \mu S + \mu T) - Q$$

et :

$$\sigma_c^2(2 Q') = \sigma_c^2(2 \Delta Q) = M_c(2 Q') [1 - M_c(2 Q')] / N$$

d'où :

$$\sigma_c^2(Q') = M_c(Q') [1 - M_c(2Q')]/2N = p(1-p) [1 - 2p(1-p)]/2N + 0(\mu/N)$$

Cette variance est égale à la variance conditionnée de ΔQ correspondant à $\Delta t = 1$; plus généralement on a :

$$\sigma_c^2(\Delta Q) = p(1-p) [1 - 2p(1-p)] \Delta t/2N$$

et :

$$M_c(\Delta Q) = [p(1-p)(1 + \mu S + \mu T) - Q] \Delta t$$

et par conséquent :

$$\begin{aligned} M_c[(\Delta Q)^2] &= \sigma_c^2(\Delta Q) + [M_c(\Delta Q)]^2 \\ &= p(1-p) [1 - 2p(1-p)] \Delta t/2N + 0(\Delta t)^2 + 0(\mu/N) \end{aligned}$$

De même pour $\Delta t = 1$, le moment centré conditionné de $(\Delta P \cdot \Delta Q)$ a pour expression :

$$\mu_c(\Delta P \cdot \Delta Q) = \mu_c(P' \cdot 2Q') = -M_c(P') \cdot M_c(2Q')/N$$

On a donc d'une façon plus générale :

$$\mu_c(\Delta P \cdot \Delta Q) = -M_c(P') \cdot M_c(Q') \cdot \Delta t/N = -p^2 \cdot p(1-p) \Delta t/N + 0(\mu/N)$$

Or :

$$\mu_c(\Delta P \cdot \Delta Q) = M_c(\Delta P \cdot \Delta Q) - M_c(\Delta P) \cdot M_c(\Delta Q)$$

Donc :

$$M_c(\Delta P \cdot \Delta Q) = -p^3(1-p) \Delta t/N + 0(\Delta t)^2 + 0(\mu/N)$$

Finalement :

$$\begin{aligned} M_c[(\Delta p)^2] &= M_c[(\Delta P + \Delta Q)^2] = M_c[(\Delta P)^2] + M_c[(\Delta Q)^2] + 2M_c(\Delta P \cdot \Delta Q) \\ &= \Delta t \cdot \left\{ p^2(1-p^2)/N + p(1-p) [1 - 2p(1-p)]/2N - 2p^3(1-p)/N \right\} \\ &\quad + 0(\Delta t)^2 + 0(\mu/N) \end{aligned}$$

Et avec l'approximation adoptée, on trouve, toutes réductions faites :

$$(8) \quad B_{20}^2 \Delta t = M_c(\Delta p)^2 = p(1-p) \Delta t/2N$$

Calculons enfin B_{11}^2 :

$$\begin{aligned} M_c(\Delta p \cdot \Delta P) &= M_c[(\Delta P)^2 + \Delta P \cdot \Delta Q] = M_c[(\Delta P)^2] + M_c(\Delta P \cdot \Delta Q) \\ &= \Delta t \cdot [p^2(1-p^2)/N - p^3(1-p)/N] \end{aligned}$$

d'où finalement :

$$(9) \quad B_{11}^2 \Delta t = p^2(1-p) \Delta t/N$$

Les autres termes figurant dans l'équation (5) sont par rapport à $1/N$ d'ordre au moins égal à 2. Pour pouvoir écrire l'équation (5) il nous suffit donc, avec l'approximation convenue, de connaître l'expression de la quantité T , qui intervient dans les formules (6), en fonction de p et de P .

Or, en désignant par H_{AA} , H_{Aa} , H_{aa} les valeurs du caractère secondaire relatives aux différents génotypes, I , J , K représentant toujours les valeurs centrées de ce même caractère, on a :

$$\begin{aligned} pT &= PI + QJ = PH_{AA} + QH_{Aa} - p \cdot M(H) \\ &= PH_{AA} + QH_{Aa} - p(PH_{AA} + 2QH_{Aa} + RH_{aa}) \end{aligned}$$

et enfin :

$$(10) \quad pT = (H_{AA} - H_{Aa})P(1-p) + (H_{Aa} - H_{aa})p(1+P-2p)$$

Reprenons l'hypothèse faite dans la première partie, chapitres II et III, d'un caractère primaire dépendant d'un seul gène que nous supposons sans dominance. On a alors, en utilisant les mêmes notations que dans les chapitres indiqués :

$$H_{AA} - H_{Aa} = -(2\alpha - \underline{O})^2 + (\alpha - \underline{O})^2 = \alpha(2\underline{O} - 3\alpha)$$

$$H_{Aa} - H_{aa} = -(\alpha - \underline{O})^2 + \underline{O}^2 = \alpha(2\underline{O} - \alpha)$$

d'où :

$$pT = \alpha [P(2\underline{O} - 3\alpha) + p(2\underline{O} - \alpha) + pP \cdot 2\alpha + 2p^2(\alpha - 2\underline{O})]$$

Finalement, les expressions (6) s'écrivent :

$$(11) \quad \begin{aligned} A_{10} &= \mu\alpha [P(2\underline{O} - 3\alpha) + p(2\underline{O} - \alpha) + 2pP\alpha - 2p^2(2\underline{O} - \alpha)] \\ A_{01} &= p^2 - P + 2\mu\alpha p [P(2\underline{O} - 3\alpha) + p(2\underline{O} - \alpha) + 2pP\alpha - 2p^2(2\underline{O} - \alpha)] \end{aligned}$$

En substituant aux quantités A_{10} , A_{01} , B_{20} ,... les expressions (11), (7), (8) et (9), dans l'équation (5), on obtient l'équation suivante :

$$(12) \quad \left\{ \begin{aligned} \frac{\partial F}{\partial t} &= x\alpha\mu \left[(2\underline{O} - 3\alpha) \frac{\partial F}{\partial y} + (2\underline{O} - \alpha) \frac{\partial F}{\partial x} + 2\alpha \frac{\partial^2 F}{\partial x \partial y} - 2(2\underline{O} - \alpha) \frac{\partial^2 F}{\partial x^2} \right] \\ &+ y \left\{ \frac{\partial^2 F}{\partial x^2} - \frac{\partial F}{\partial y} + 2\mu\alpha \left[(2\underline{O} - 3\alpha) \frac{\partial^2 F}{\partial x \partial y} + (2\underline{O} - \alpha) \frac{\partial^2 F}{\partial x^2} + 2\alpha \frac{\partial^3 F}{\partial x^2 \partial y} - 2(2\underline{O} - \alpha) \frac{\partial^3 F}{\partial x^3} \right] \right\} \\ &- \left(\frac{\partial^2 F}{\partial x^2} - \frac{\partial F}{\partial x} \right) x^2/4N - \left(\frac{\partial^3 F}{\partial x^3} - \frac{\partial^2 F}{\partial x^2} \right) xy/N - \left(\frac{\partial^4 F}{\partial x^4} - \frac{\partial^2 F}{\partial x^2} \right) y^2/2N \end{aligned} \right\}$$

équation aux dérivées partielles linéaires du quatrième ordre.

* * *

Remarquons que si l'on donne à y la valeur 0, la fonction $F(x, 0, t)$ n'est autre que la fonction caractéristique à l'instant t de la seule variable aléatoire p ⁽¹⁾. L'équation (12) devient pour $y = 0$:

$$(13) \quad \frac{\partial F(x, 0, t)}{\partial t} = x^{\alpha\mu} \left[(2\underline{0} - 3\alpha) \frac{\partial F(x, 0, t)}{\partial y} + (2\underline{0} - \alpha) \frac{\partial F(x, 0, t)}{\partial x} + 2\alpha \frac{\partial^2 F(x, 0, t)}{\partial x \partial y} - 2(2\underline{0} - \alpha) \frac{\partial^2 F(x, 0, t)}{\partial x^2} \right] - \left[\frac{\partial^2 F(x, 0, t)}{\partial x^2} - \frac{\partial F(x, 0, t)}{\partial x} \right] x^2/4N$$

Mais, si l'on tient compte du fait que la fonction $F(x, y, t)$ est une fonction caractéristique, on peut exprimer simplement les dérivées partielles :

$$\frac{\partial F(x, 0, t)}{\partial y} \quad \text{et} \quad \frac{\partial^2 F(x, 0, t)}{\partial x \partial y}$$

en fonction des dérivées partielles relatives à la seule variable x .

En effet, on a :

$$\begin{aligned} \frac{\partial F(x, y, t)}{\partial y} &= M(P \cdot e^{px} + P_y) & \text{donc :} & \quad \frac{\partial F(x, 0, t)}{\partial y} = M(P \cdot e^{px}) \\ \frac{\partial F(x, y, t)}{\partial x} &= M(p \cdot e^{px} + P_y) & \text{donc :} & \quad \frac{\partial F(x, 0, t)}{\partial x} = M(p \cdot e^{px}) \end{aligned}$$

et, de la même manière, on trouve :

$$\frac{\partial^2 F(x, 0, t)}{\partial x^2} = M(p^2 e^{px}); \quad \frac{\partial^2 F(x, 0, t)}{\partial x \partial y} = M(p P e^{px}); \quad \frac{\partial^3 F(x, 0, t)}{\partial x^3} = M(p^3 e^{px})$$

Proposons-nous de calculer $M(P \cdot e^{px})$; supposons connue la génération F_n , désignons toujours par M_c les moyennes conditionnées lorsque p et P sont connues et désignons de plus par M_c la moyenne conditionnée lorsque en outre p' est donnée. On a pour la génération F_{n+1} :

$$M_c(P' e^{p'x}) = M_c[e^{p'x} \cdot M_c(P')]$$

(1) En posant $x = LX$, on obtient la « fonction génératrice » $M(X^p)$ de p .

Or cette moyenne conditionnée $M_c(P')$ peut s'exprimer en utilisant le schéma suivant :

Les N individus de la génération F_{n+1} sont choisis au hasard parmi tous les enfants possibles issus de la génération F_n ; l'ensemble de ces enfants présente les génotypes AA , Aa , aa , avec les probabilités respectives :

$$p^2 (1 + 2 \mu T) ; \quad 2 pq (1 + \mu T + \mu S) ; \quad q^2 (1 + 2 \mu S)$$

Classons ces enfants en trois groupes suivant leur génotype. D'autre part, répartissons les individus de F_{n+1} en deux catégories, la première contenant les homozygotes AA et la moitié des hétérozygotes, la deuxième contenant les homozygotes aa et la moitié des hétérozygotes. La fréquence des individus de la première catégorie est p' . Parmi ceux-ci, la probabilité pour qu'un individu pris au hasard soit un homozygote AA , c'est-à-dire provienne du premier groupe des enfants possibles de F_n est :

$$\frac{p^2 (1 + 2 \mu T)}{p^2 (1 + 2 \mu T) + pq (1 + \mu T + \mu S)} = \frac{p^2 (1 + 2 \mu T)}{p (1 + \mu T)} = p + 0 (\mu)$$

La moyenne conditionnée $M_c(P')$ est donc :

$$M_c(P') = p' \cdot p + 0 (\mu)$$

On en déduit :

$$M_c(P'e^{p'x}) = p \cdot M_c(p' \cdot e^{p'x}) + 0 (\mu)$$

et par conséquent :

$$(14) \quad M(P'e^{p'x}) = M[(P + \Delta P)e^{(p + \Delta p)x}] = M\left\{p \cdot M_c[(p + \Delta p)e^{(p + \Delta p)x}]\right\} + 0 (\mu)$$

Or, on a :

$$\begin{aligned} M[(P + \Delta P)e^{(p + \Delta p)x}] &= M(Pe^{px} \cdot e^{x\Delta p} + \Delta P e^{px} \cdot e^{x\Delta p}) \\ &= M\left\{Pe^{px} \cdot M_c[1 + x\Delta p + x^2(\Delta p)^2/2 + \dots] + e^{px} M_c[\Delta P + x\Delta P\Delta p + x^2\Delta P(\Delta p)^2/2 + \dots]\right\} \\ &= M(Pe^{px}) + 0 (\mu) + 0 (1/N) \end{aligned}$$

et d'autre part :

$$\begin{aligned} M\left\{p \cdot M_c[(p + \Delta p)e^{(p + \Delta p)x}]\right\} &= M\left[p^2 e^{px} M_c(e^{x\Delta p}) + p e^{px} M_c(\Delta p e^{x\Delta p})\right] \\ &= M\left\{p^2 e^{px} M_c[1 + x\Delta p + x^2(\Delta p)^2/2 + \dots] + p e^{px} M_c[\Delta p + x(\Delta p)^2 + \dots]\right\} \\ &= M(p^2 e^{px}) + 0 (\mu) + 0 (1/N) \end{aligned}$$

d'où, d'après (14) :

$$M(Pe^{px}) = M(p^2 e^{px}) + 0(\mu) + 0(1/N)$$

On a donc entre les dérivées partielles de $F(x, 0, t)$ la relation :

$$(15) \quad \frac{\partial F(x, 0, t)}{\partial y} = \frac{\partial^2 F(x, 0, t)}{\partial x^2} + 0(\mu) + 0(1/N)$$

On a de même :

$$(15') \quad \frac{\partial^2 F(x, 0, t)}{\partial x \partial y} = \frac{\partial^3 F(x, 0, t)}{\partial x^3} + 0(\mu) + 0(1/N)$$

Et finalement, en désignant simplement par $F(x, t)$ la fonction caractéristique de la seule variable aléatoire p , cette fonction satisfait à l'équation aux dérivées partielles suivante, déduite de l'équation (13) en négligeant, suivant l'approximation adoptée, $0(\mu^2)$, $0(1/N^2)$, $0(\mu/N)$:

$$\frac{\partial F}{\partial t} = x\alpha\mu \left[(2\underline{0} - 3\alpha) \frac{\partial^2 F}{\partial x^3} + (2\underline{0} - \alpha) \frac{\partial F}{\partial x} + 2\alpha \frac{\partial^3 F}{\partial x^3} - 2(2\underline{0} - \alpha) \frac{\partial^2 F}{\partial x^2} \right] + \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) x^2 / 4N$$

équation linéaire du troisième ordre, que nous désignerons dorénavant par « équation (E) » et qui peut s'écrire :

$$(E) \quad \frac{\partial F}{\partial t} = \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) [(2\underline{0} - \alpha)x\alpha\mu + x^2/4N] - \left(\frac{\partial^2 F}{\partial x^2} - \frac{\partial^3 F}{\partial x^3} \right) \cdot 2x\alpha^2\mu$$

*
* * *

Remarquons tout d'abord que pour $t = 0$, c'est-à-dire pour la population initiale, p est déterminée et égale à p_0 , donc la fonction caractéristique $F(x, t)$ doit satisfaire à la condition :

$$(16) \quad F(x, 0) = e^{p_0 x}$$

Par ailleurs on a :

$$\frac{\partial^2 F}{\partial x^2} - \frac{\partial^3 F}{\partial x^3} = M \left[(p^2 - p^3) e^{px} \right] = M \left[(1 - p) p^2 e^{px} \right]$$

$$\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} = M \left[(p - p^2) e^{px} \right] = M \left[(1 - p) p e^{px} \right]$$

On en déduit puisque $p^2 \leq p \leq 1$:

$$0 \leq \frac{\partial^2 F}{\partial x^2} - \frac{\partial^3 F}{\partial x^3} \leq \frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2}$$

et donc, si $x \geq 0$, les inégalités suivantes :

$$(17) \quad \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) [(2\underline{O} - 3\alpha)x^{\alpha\mu} + x^2/4N] \leq \frac{\partial F}{\partial t} \leq \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) [(2\underline{O} - \alpha)x^{\alpha\mu} + x^2/4N]$$

ou les inégalités contraires si $x \leq 0$.

Ces inégalités montrent que $\frac{\partial F}{\partial t}$ est nulle ou a un signe constant pour toute valeur

déterminée x_0 de x extérieure à l'intervalle :

$$[4N(\alpha - 2\underline{O})\alpha\mu; \quad 4N(3\alpha - 2\underline{O})\alpha\mu]$$

car les deux membres extrêmes sont alors nuls ou de même signe. Il en résulte que pour une telle valeur x_0 , la fonction de t : $F(x_0, t)$ est monotone et bornée (on a en effet $1 \leq F \leq e^{x_0}$, ou les inégalités contraires suivant le signe de x_0) et par conséquent tend vers une limite finie $\varphi(x_0)$ quand t augmente indéfiniment.

Mais ce raisonnement ne suffit pas pour que nous puissions affirmer l'existence d'une loi limite de probabilité dont la fonction caractéristique serait $\varphi(x)$; il nous faut encore démontrer que $F(x, t)$ d'une part a encore une limite finie lorsque :

$$4N(\alpha - 2\underline{O})\alpha\mu \leq x \leq 4N(3\alpha - 2\underline{O})\alpha\mu$$

et d'autre part tend *uniformément* vers sa limite $\varphi(x)$.

*
* *

Soit X une valeur de x supérieure à $4N(3\alpha - 2\underline{O})\alpha\mu$; nous venons de voir que $F(X, t)$ a lorsque t augmente indéfiniment la limite finie $\varphi(X)$. Donc la dérivée partielle $\frac{\partial F(X, t)}{\partial t}$ a pour limite 0 lorsque t augmente indéfiniment ; les deux membres extrêmes de (17) étant alors tous deux de même signe ou nuls, ceci n'est possible que si l'un des deux membres au moins a pour limite 0, ce qui exige que la fonction de t :

$$\frac{\partial F(X, t)}{\partial x} - \frac{\partial^2 F(X, t)}{\partial x^2}$$

tende vers 0 lorsque t augmente indéfiniment.

Or, pour une valeur donnée de t , la fonction de x

$$\frac{\partial F(x, t)}{\partial x} - \frac{\partial^2 F(x, t)}{\partial x^2}$$

est positive ou nulle et non décroissante, car on a :

$$0 \leq M \left[(1 - p) p e^{px_0} \right] \leq M \left[(1 - p) p e^{px_1} \right] \quad \text{si} \quad x_0 < x_1$$

Il en résulte que la fonction $\frac{\partial F(x, t)}{\partial x} - \frac{\partial^2 F(x, t)}{\partial x^2}$ tend *uniformément* vers 0 lorsque t augmente indéfiniment, pour $x < X$. Donc, d'après (17), $\frac{\partial F}{\partial t}$ tend vers 0 pour toute valeur de x .

Les inégalités (17) montrent d'autre part que l'on a toujours au moins l'une des deux inégalités suivantes :

$$(18) \quad \left| \frac{\partial F}{\partial t} \right| \leq \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) \left| (2 \underline{Q} - \alpha) x^{\alpha \mu} + x^2/4 N \right|$$

$$\left| \frac{\partial F}{\partial t} \right| \leq \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) \left| (2 \underline{Q} - 3 \alpha) x^{\alpha \mu} + x^2/4 N \right|$$

Considérons alors un intervalle quelconque dont X (que nous supposons > 0) soit la borne supérieure, et désignons par M le plus grand des maxima absolus des deux fonctions de x :

$$\left| (2 \underline{Q} - \alpha) x^{\alpha \mu} + x^2/4 N \right| \quad \text{et} \quad \left| (2 \underline{Q} - 3 \alpha) x^{\alpha \mu} + x^2/4 N \right|$$

dans cet intervalle ; on a d'après (18) :

$$(19) \quad \left| \frac{\partial F}{\partial t} \right| \leq \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) M$$

pour toute valeur x de l'intervalle ; ce qui prouve que $\frac{\partial F}{\partial t}$ converge *uniformément* vers 0 dans l'intervalle considéré puisque le deuxième membre de (20) converge uniformément vers 0 pour toute valeur de x inférieure à X .

Montrons alors que dans cet intervalle la fonction $F(x, t)$ converge uniformément vers une limite $\varphi(x)$ lorsque t augmente indéfiniment.

Pour cela, il suffit de montrer que l'on peut déterminer T tel que pour $t > T$ on ait, quel que soit x dans l'intervalle considéré :

$$(20) \quad \left| \int_t^{\infty} \frac{\partial F(x, t)}{\partial t} dt \right| < \varepsilon$$

Or, on a d'après ce qui précède :

$$\left| \frac{\partial F(x, t)}{\partial t} \right| \leq \left[\frac{\partial F(x, t)}{\partial x} - \frac{\partial^2 F(x, t)}{\partial x^2} \right] M \leq \left[\frac{\partial F(X, t)}{\partial x} - \frac{\partial^2 F(X, t)}{\partial x^2} \right] M$$

et d'après (17) puisque X est supposé > 0 :

$$\frac{\partial F(X, t)}{\partial x} - \frac{\partial^2 F(X, t)}{\partial x^2} \leq \frac{\partial F(X, t)}{\partial t} \cdot \frac{1}{(2 \underline{Q} - 3 \alpha) X_{\alpha\mu} + X^2/4 N}$$

Donc :

$$\left| \int_t^{\infty} \frac{\partial F(x, t)}{\partial t} dt \right| \leq \int_t^{\infty} \left| \frac{\partial F(x, t)}{\partial t} \right| dt \leq \int_t^{\infty} \frac{\partial F(X, t)}{\partial t} \cdot \frac{M dt}{(2 \underline{Q} - 3 \alpha) X_{\alpha\mu} + X^2/4 N}$$

Or :

$$\int_t^{\infty} \frac{\partial F(X, t)}{\partial t} dt = \varphi(X) - F(X, t)$$

et cette quantité tendant vers 0 lorsque t augmente indéfiniment il suffit pour réaliser la condition (20) de prendre pour T une valeur telle que :

$$\left| \varphi(X) - F(X, T) \right| < \varepsilon \cdot \frac{(2 \underline{Q} - 3 \alpha) X_{\alpha\mu} + X^2/4 N}{M}$$

ce qui est toujours possible.

En résumé donc, avec l'approximation adoptée, la fonction caractéristique $F(x, t)$ converge uniformément vers une limite $\varphi(x)$ dans tout intervalle fini. Cette limite $\varphi(x)$ est donc la fonction caractéristique de la loi limite de probabilité de la fréquence p .

Forme de la loi limite.

Si les dérivées $\frac{\partial F}{\partial x}$ et $\frac{\partial^2 F}{\partial x^2}$ tendent uniformément vers des limites respectives $\varphi_1(x)$ et $\varphi_2(x)$, on a nécessairement :

$$\varphi_1(x) = \frac{d\varphi}{dx} \quad \text{et} \quad \varphi_2(x) = \frac{d\varphi_1(x)}{dx} = \frac{d^2\varphi(x)}{dx^2}$$

En effet, supposons par exemple que $\frac{\partial F}{\partial x}$ tende uniformément vers $\varphi_1(x)$; soit X donné > 0 , on peut déterminer T_1 tel que pour $0 < x < X$ et $t > T_1$ on ait :

$$\left| \int_0^x \left[\frac{\partial F}{\partial x} - \varphi_1(x) \right] dx \right| = \left| F(x) - F(0) - \int_0^x \varphi_1(x) dx \right| < \varepsilon_1 X = \varepsilon/3$$

ε_1 et donc ε étant des nombres > 0 arbitrairement petits.

On peut d'autre part déterminer T_2 de telle sorte que pour $0 < x < X$ et $t > T_2$ on ait :

$$\left| F(0) - \varphi(0) \right| < \varepsilon/3 \quad \text{et} \quad \left| F(x) - \varphi(x) \right| < \varepsilon/3$$

Donc, pour $t > \begin{Bmatrix} T_1 \\ T_2 \end{Bmatrix}$ on a :

$$\left| F(x) - F(0) + F(0) - \varphi(0) - F(x) + \varphi(x) - \int_0^x \varphi_1(x) dx \right| < \varepsilon$$

ou :

$$\left| \varphi(x) - \varphi(0) - \int_0^x \varphi_1(x) dx \right| < \varepsilon$$

ou enfin :

$$\left| \int_0^x \left[\frac{d\varphi}{dx} - \varphi_1(x) \right] dx \right| < \varepsilon$$

Cette inégalité étant vérifiée quel que soit x ($x < X$) et ε étant arbitrairement petit, ceci exige que l'on ait :

$$\frac{d\varphi}{dx} = \varphi_1(x)$$

Montrons donc que la dérivée $\frac{\partial F}{\partial x}$ converge uniformément vers une fonction $\varphi_1(x)$ lorsque t augmente indéfiniment.

En dérivant (E) par rapport à x on obtient :

$$(E_1) \quad \frac{\partial}{\partial t} \left(\frac{\partial F}{\partial x} \right) = \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) \left[(2 \underline{0} - \alpha) \alpha \mu + \frac{x}{2N} \right] + \left(\frac{\partial^2 F}{\partial x^2} - \frac{\partial^3 F}{\partial x^3} \right) \left[\frac{x^2}{4N} + (2 \underline{0} - \alpha) x \alpha \mu - 2 \alpha^2 \mu \right] - \left(\frac{\partial^3 F}{\partial x^3} - \frac{\partial^4 F}{\partial x^4} \right) 2 x \alpha^2 \mu$$

En tenant compte de :

$$0 \leq \frac{\partial^3 F}{\partial x^3} - \frac{\partial^4 F}{\partial x^4} \leq \frac{\partial^2 F}{\partial x^2} - \frac{\partial^3 F}{\partial x^3} \leq \frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2}$$

on a si x est > 0 :

$$(17') \quad \begin{cases} \frac{\partial}{\partial t} \left(\frac{\partial F}{\partial x} \right) \leq \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) \left[\frac{x}{2N} + (2 \underline{0} - \alpha) \alpha \mu \right] + \left(\frac{\partial^2 F}{\partial x^2} - \frac{\partial^3 F}{\partial x^3} \right) \left[\frac{x^2}{4N} + (2 \underline{0} - \alpha) x \alpha \mu - 2 \alpha^2 \mu \right] \\ \frac{\partial}{\partial t} \left(\frac{\partial F}{\partial x} \right) \geq \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) \left[\frac{x}{2N} + (2 \underline{0} - \alpha) \alpha \mu \right] + \left(\frac{\partial^2 F}{\partial x^2} - \frac{\partial^3 F}{\partial x^3} \right) \left[\frac{x^2}{4N} + (2 \underline{0} - 3\alpha) x \alpha \mu - 2 \alpha^2 \mu \right] \end{cases}$$

et les inégalités contraires si x est < 0 .

Pour $x = X$, assez grand et > 0 , les deuxièmes membres des inégalités (17') sont tous deux > 0 ; $\frac{\partial}{\partial t} \left(\frac{\partial F}{\partial x} \right)$ est alors > 0 ; donc $\frac{\partial F}{\partial X}$, fonction de t non décroissante et bornée (on a en effet $0 \leq \frac{\partial F}{\partial X} \leq e^x$) tend vers une limite $\varphi_1(X)$.

D'autre part, les inégalités (17') montrent que l'on a quel que soit x au moins l'une des inégalités suivantes :

$$(18') \quad \begin{cases} \left| \frac{\partial}{\partial t} \left(\frac{\partial F}{\partial x} \right) \right| \leq \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) \left[\left| \frac{x}{2N} + (2 \underline{0} - \alpha) \alpha \mu \right| + \left| \frac{x^2}{4N} + (2 \underline{0} - 3\alpha) x \alpha \mu - 2 \alpha^2 \mu \right| \right] \\ \left| \frac{\partial}{\partial t} \left(\frac{\partial F}{\partial x} \right) \right| \leq \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) \left[\left| \frac{x}{2N} + (2 \underline{0} - \alpha) \alpha \mu \right| + \left| \frac{x^2}{4N} + (2 \underline{0} - \alpha) x \alpha \mu - 2 \alpha^2 \mu \right| \right] \end{cases}$$

Soit alors un intervalle quelconque de borne supérieure X et soit M_1 le plus grand des maxima absolus dans cet intervalle des deux fonctions de x :

$$\begin{aligned} & \left| \frac{x}{2N} + (2 \underline{0} - \alpha) \alpha \mu \right| + \left| \frac{x^2}{4N} + (2 \underline{0} - 3\alpha) x \alpha \mu - 2 \alpha^2 \mu \right| \\ \text{et} & \left| \frac{x}{2N} + (2 \underline{0} - \alpha) \alpha \mu \right| + \left| \frac{x^2}{4N} + (2 \underline{0} - \alpha) x \alpha \mu - 2 \alpha^2 \mu \right| \end{aligned}$$

on a :

(19')

$$\left| \frac{\partial}{\partial t} \left(\frac{\partial F}{\partial x} \right) \right| \leq \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) M_1$$

Il suffit de montrer que l'on peut déterminer T de telle sorte que, pour $t > T$, on ait quel que soit x dans l'intervalle considéré :

(20')

$$\left| \int_t^\infty \frac{\partial}{\partial t} \left[\frac{\partial F(x, t)}{\partial x} \right] dt \right| < \varepsilon$$

Or, d'après (19') on a

$$\left| \frac{\partial}{\partial t} \left[\frac{\partial F(x, t)}{\partial x} \right] \right| \leq \left[\frac{\partial F(x, t)}{\partial x} - \frac{\partial^2 F(x, t)}{\partial x^2} \right] M_1 \leq \left[\frac{\partial F(X, t)}{\partial X} - \frac{\partial^2 F(X, t)}{\partial X^2} \right] M_1$$

Mais, X étant > 0 assez grand, les inégalités (17) page 96 permettent d'écrire :

$$\frac{\partial F(X, t)}{\partial X} - \frac{\partial^2 F(X, t)}{\partial X^2} \leq \frac{\partial F(X, t)}{\partial t} \cdot \frac{1}{X^2/4N + (2\underline{O} - 3\alpha) X_{\alpha\mu}}$$

On a par conséquent :

$$\left| \int_t^\infty \frac{\partial}{\partial t} \left(\frac{\partial F}{\partial x} \right) dt \right| \leq \int_t^\infty \left| \frac{\partial}{\partial t} \left(\frac{\partial F}{\partial x} \right) \right| dt \leq \int_t^\infty \frac{\partial F(X, t)}{\partial t} \cdot \frac{M_1 dt}{X^2/4N + (2\underline{O} - 3\alpha) X_{\alpha\mu}}$$

et il suffit pour réaliser (20') de prendre pour T une valeur telle que pour $t > T$ on ait :

$$\varphi(X) - F(X, t) < \varepsilon \frac{(2\underline{O} - 3\alpha) X_{\alpha\mu} + X^2/4N}{M_1}$$

ce qui est toujours possible.

On peut démontrer d'une façon analogue que la dérivée $\frac{\partial^2 F}{\partial x^2}$ converge uniformément vers une fonction limite.

En effet, dérivons (E₁) par rapport à x , on obtient :

(E₂)

$$\begin{aligned} \frac{\partial}{\partial t} \left(\frac{\partial^2 F}{\partial x^2} \right) &= \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) / 2N + \left(\frac{\partial^2 F}{\partial x^2} - \frac{\partial^3 F}{\partial x^3} \right) \left[x/N + (2\underline{O} - \alpha) 2\alpha\mu \right] \\ &+ \left(\frac{\partial^3 F}{\partial x^3} - \frac{\partial^4 F}{\partial x^4} \right) \left[x^2/4N + (2\underline{O} - \alpha) x\alpha\mu - 4\alpha^2\mu \right] - \left(\frac{\partial^4 F}{\partial x^4} - \frac{\partial^5 F}{\partial x^5} \right) 2x\alpha^2\mu \end{aligned}$$

En tenant compte de :

$$0 \leq \frac{\partial^4 F}{\partial x^4} - \frac{\partial^5 F}{\partial x^5} \leq \frac{\partial^3 F}{\partial x^3} - \frac{\partial^4 F}{\partial x^4}$$

on a, si x est > 0 , les inégalités suivantes :

$$(17'') \quad \begin{aligned} \frac{\partial}{\partial t} \left(\frac{\partial^2 F}{\partial x^2} \right) &\geq \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) / 2N + \left(\frac{\partial^2 F}{\partial x^2} - \frac{\partial^3 F}{\partial x^3} \right) \left[x/N + (2\underline{O} - \alpha) 2\alpha\mu \right] + \left(\frac{\partial^3 F}{\partial x^3} - \frac{\partial^4 F}{\partial x^4} \right) \\ &\quad [x^2/4N + (2\underline{O} - \alpha) x\alpha\mu - 6\alpha^2\mu] \\ \frac{\partial}{\partial t} \left(\frac{\partial^2 F}{\partial x^2} \right) &\leq \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) / 2N + \left(\frac{\partial^2 F}{\partial x^2} - \frac{\partial^3 F}{\partial x^3} \right) \left[x/N + (2\underline{O} - \alpha) 2\alpha\mu \right] + \left(\frac{\partial^3 F}{\partial x^3} - \frac{\partial^4 F}{\partial x^4} \right) \\ &\quad [x^2/4N + (2\underline{O} - \alpha) x\alpha\mu - 4\alpha^2\mu] \end{aligned}$$

ou les inégalités contraires si x est < 0 .

Pour $x = X > 0$ assez grand, les deux deuxièmes membres des inégalités (17'') sont > 0 , et la fonction $\frac{\partial^2 F}{\partial X^2}(X, t)$, non décroissante et bornée ($0 \leq \frac{\partial^2 F(X, t)}{\partial X^2} \leq e^X$) tend vers une limite finie lorsque t augmente indéfiniment.

Les inégalités (17'') montrent que l'on a quel que soit x au moins l'une des inégalités :

$$(18'') \quad \begin{aligned} \left| \frac{\partial}{\partial t} \left(\frac{\partial^2 F}{\partial x^2} \right) \right| &\leq \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) \left[1/2N + \left| x/N + (2\underline{O} - \alpha) 2\alpha\mu \right| + \left| x^2/4N + (2\underline{O} - \alpha) x\alpha\mu - 6\alpha^2\mu \right| \right] \\ \left| \frac{\partial}{\partial t} \left(\frac{\partial^2 F}{\partial x^2} \right) \right| &\leq \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) \left[1/2N + \left| x/N + (2\underline{O} - \alpha) 2\alpha\mu \right| + \left| x^2/4N + (2\underline{O} - \alpha) x\alpha\mu - 4\alpha^2\mu \right| \right] \end{aligned}$$

Considérons alors un intervalle quelconque de borne supérieure égale à X . On a, en désignant par M_2 le plus grand des maxima absolus, dans cet intervalle, des deux fonctions :

$$\begin{aligned} &1/2N + \left| x/N + (2\underline{O} - \alpha) 2\alpha\mu \right| + \left| x^2/4N + (2\underline{O} - \alpha) x\alpha\mu - 6\alpha^2\mu \right| \\ \text{et} \quad &1/2N + \left| x/N + (2\underline{O} - \alpha) 2\alpha\mu \right| + \left| x^2/4N + (2\underline{O} - \alpha) x\alpha\mu - 4\alpha^2\mu \right| \end{aligned}$$

l'inégalité :

$$(19'') \quad \left[\frac{\partial}{\partial t} \left(\frac{\partial^2 F}{\partial x^2} \right) \right] \leq \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) M_2$$

On en déduit par un raisonnement identique à celui que nous venons de faire pour la fonction $\frac{\partial F(x, t)}{\partial x}$, que la fonction $\frac{\partial^2 F(x, t)}{\partial x^2}$ converge uniformément vers sa fonction limite.

Nous pouvons maintenant prévoir la forme de la loi limite de probabilité de la fréquence p . En effet, nous avons vu que la fonction

$$\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2}$$

tend uniformément vers 0 lorsque t augmente indéfiniment ; il résulte par ailleurs de ce qui précède que les dérivées $\frac{\partial F}{\partial x}$ et $\frac{\partial^2 F}{\partial x^2}$ ont respectivement pour limites $\frac{d\varphi}{dx}$ et $\frac{d^2\varphi}{dx^2}$. La limite $\varphi(x)$ de $F(x, t)$ doit donc vérifier l'équation :

$$\frac{d\varphi}{dx} - \frac{d^2\varphi}{dx^2} = 0$$

D'où :

(21)

$$\varphi(x) = ae^x + b$$

où a et b sont des constantes à déterminer.

Remarquons tout d'abord que, pour $x = 0$, l'équation (E) donne

$$\frac{\partial F}{\partial t} = 0$$

donc la fonction $F(x, t)$ est stationnaire, et :

$$\varphi(0) = F(0, 0) = 1$$

d'où :

$$a + b = 1$$

La fonction $\varphi(x)$ s'écrit donc :

(22)

$$\varphi(x) = a(e^x - 1) + 1$$

C'est la fonction caractéristique d'une variable aléatoire qui ne peut prendre que les valeurs 1 (avec la probabilité a) et 0 (avec la probabilité $1 - a$). La population tend donc asymptotiquement vers l'homogénéité.

La détermination, sous certaines hypothèses complémentaires, de la valeur de a ainsi que de la rapidité avec laquelle cette homogénéité est atteinte fera l'objet des chapitres suivants.

CHAPITRE II

ÉVOLUTION D'UNE POPULATION LIMITÉE MAL ADAPTÉE

Nous supposons dans ce chapitre et dans les suivants que la quantité α , contribution du gène A au caractère primaire, est petite.

La mauvaise adaptation de la population se traduit par le fait que l'optimum $\underline{Q}$ est (en valeur absolue) grand par rapport à α .

L'équation (E) peut s'écrire :

$$\frac{\partial F}{\partial t} = \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) (x - \xi) x / 4 N - \left[\left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) + 2 \left(\frac{\partial^2 F}{\partial x^2} - \frac{\partial^3 F}{\partial x^3} \right) \right] \alpha^2 \mu x$$

en posant :

$$\xi = -8 N \alpha \underline{Q} \mu$$

Le deuxième terme du deuxième membre de (1) est petit par rapport au premier. Considérons alors comme première approximation de la fonction F la fonction G satisfaisant à l'équation :

$$\frac{\partial G}{\partial t} = \left(\frac{\partial G}{\partial x} - \frac{\partial^2 G}{\partial x^2} \right) (x - \xi) x / 4 N$$

Reprenons sur cette équation (3) le raisonnement qu'a fait M. MALÉCOT sur une équation parabolique analogue⁽¹⁾. Les quantités $G(0, t)$ et $G(\xi, t)$ sont indépendantes de t , et il résulte de la condition initiale

$$G(x, 0) = e^{p_0 x}$$

les conditions :

$$G(0, t) = 1$$

$$G(\xi, t) = e^{p_0 \xi}$$

(1) G. MALÉCOT, *Les processus stochastiques et la méthode des fonctions génératrices ou caractéristiques*, p. 9, équation (11) (Publications de l'Institut de Statistique, Paris, vol. I, fasc. III, 1952).

La solution stationnaire de l'équation (3) qui satisfait aux conditions (5) a pour expression

$$(6) \quad g(x) = \frac{1 - e^{p_0 \xi}}{1 - e^{\xi}} (e^x - 1) + 1$$

Posons :

$$G(x, t) - g(x) = Z(x, t)$$

Z est une solution de (3) qui s'annule pour $x = 0$ et $x = \xi$. C'est donc une série de Sturm Liouville de la forme :

$$Z(x, t) = \sum_{i=1}^{\infty} k_i f_i(x) h_i(t)$$

chaque produit $f_i h_i$ vérifiant (3) et f_i étant nulle pour $x = 0$ et $x = \xi$; les constantes k_i étant déterminées par la condition (4).

Les fonctions f et h satisfont à l'équation :

$$\frac{h'}{h} = \left(\frac{f'}{f} - \frac{f''}{f} \right) (x - \xi) x/4 N = -s$$

s étant une constante.

D'où :

$$h_i(t) = e^{-s_i t}$$

et :

$$(7) \quad (f_i' - f_i'')(x - \xi) x/4 N + s_i f_i = 0$$

les constantes s_i étant les valeurs propres relatives à l'intervalle $(0, \xi)$ de l'équation différentielle (7). Et on a finalement :

$$(8) \quad G(x, t) = \frac{1 - e^{p_0 \xi}}{1 - e^{\xi}} (e^x - 1) + 1 + \sum_{i=1}^{\infty} k_i e^{-s_i t} f_i(x)$$

Il est nécessaire de préciser ce résultat avant d'utiliser (8) pour la recherche d'une meilleure approximation de la fonction $F(x, t)$ et de son expression asymptotique.

Détermination des valeurs propres s_i et des fonctions f_i :

Transformons l'équation (7) en prenant comme nouvelle inconnue

$$X = x/\xi$$

On a immédiatement :

$$X(X-1) \frac{d^2 f}{dX^2} - \xi X(X-1) \frac{df}{dX} - 4 N_s f = 0$$

Nous sommes donc amenés (en changeant de notations) à chercher les solutions de l'équation différentielle :

(9)

$$x(x-1)y'' - \xi x(x-1)y' - 4N_s y = 0$$

qui s'annulent pour $x = 0$ et $x = 1$, et à déterminer les valeurs propres s .

Or, l'équation (9) peut s'écrire :

(10)

$$x^2 y'' + xy' + A(x)y + B(x)y = 0$$

avec :

$$A(x) = -\xi x; \quad B(x) = -\frac{4N_s x}{x-1}$$

$A(x)$ et $B(x)$ étant holomorphes dans le domaine de l'origine, l'équation (9) est *du type de Fuchs* et a deux intégrales régulières distinctes dans le domaine de l'origine.

En posant :

$$y = x^r + c_1 x^{r+1} + \dots + c_n x^{r+n} + \dots$$

on obtient, après substitution dans (10), en annulant le coefficient de x^r , l'équation déterminante :

$$r(r-1) = 0$$

Elle a deux racines 1 et 0 ; comme celles-ci diffèrent d'un nombre entier, les deux intégrales considérées sont de la forme :

$$y_1 = x \cdot \Psi(x)$$

$$y_2 = x \cdot \Psi(x) [\text{CLog} x + \chi(x)/x]$$

les fonctions $\Psi(x)$ et $\chi(x)$ étant holomorphes dans le domaine de l'origine et ne s'annulant pas pour $x = 0$. Il en résulte que y_2 ne s'annule pas pour $x = 0$ et donc que l'intégrale générale $\lambda_1 y_1 + \lambda_2 y_2$ ne peut s'annuler pour $x = 0$ que si $\lambda_2 = 0$.

Les intégrales cherchées sont donc de la forme

$$y = \lambda y_1$$

Précisons par suite cette intégrale y_1 .

Posons :

$$y_1 = x(u_0 + u_1 x + u_2 x^2 + \dots + u_n x^n + \dots)$$

d'où :

$$y_1' = u_0 + 2 u_1 x + \dots + n u_{n-1} x^{n-1} + \dots$$

$$y_1'' = 2 u_1 + 6 u_2 x + \dots + n(n-1) u_{n-1} x^{n-2} + \dots$$

En substituant ces valeurs dans (9) et en annulant le coefficient de x^n , on obtient la relation de récurrence :

(11)

$$n(n+1) u_n = [n(n-1) + \xi] - 4 N_s] u_{n-1} - \xi(n-1) u_{n-2}$$

qui permet de calculer tous les coefficients u_i à partir des deux premiers u_0 et u_1 . Or on peut prendre arbitrairement :

$$u_0 = 1$$

et le coefficient u_1 s'obtient immédiatement en annulant le coefficient de x après substitution dans (9) des expressions de y_1 , y_1' et y_1'' ; on trouve :

$$u_1 = (\xi - 4 N_s)/2$$

Calculons alors une valeur approchée de u_n en tenant compte de ce que ξ est petit par hypothèse. Posons :

$$u_n = A_n + B_n \xi + O(\xi^2)$$

On déduit immédiatement de (11) :

$$n(n+1) A_n = [n(n-1) - 4 N_s] A_{n-1}$$

ou :

(12)

$$\frac{A_{n+1}}{A_n} = \frac{n(n+1) - 4 N_s}{(n+1)(n+2)}$$

ce qui montre que $A^n x^n$ est le terme général de la série hypergéométrique de somme :

$$\mathcal{F}(\beta, \gamma, 2, x)$$

en posant :

$$(13) \quad n(n+1) - 4Ns = (n+\beta)(n+\gamma)$$

De même la relation (11) donne :

$$(14) \quad n(n+1)B_n = [n(n-1) - 4ns]B_{n-1} + nA_{n-1} - (n-1)A_{n-2}$$

Or on peut écrire :

$$\frac{u_n}{u_{n-1}} = \frac{A_n + B_n \xi}{A_{n-1} + B_{n-1} \xi} = \frac{A_n}{A_{n-1}} + \left(\frac{B_n}{A_{n-1}} - \frac{A_n B_{n-1}}{A_{n-1}^2} \right) \xi + O(\xi^2)$$

Mais on déduit de (14) :

$$\frac{B_n}{A_{n-1}} = \frac{n(n-1) - 4Ns}{n(n+1)} \cdot \frac{B_{n-1}}{A_{n-1}} + \frac{1}{n+1} - \frac{A_{n-2}}{A_{n-1}} \cdot \frac{n-1}{n(n+1)}$$

d'où, en tenant compte de (12) :

$$\frac{B_n}{A_{n-1}} - \frac{A_n B_{n-1}}{A_{n-1}^2} = \frac{1}{n+1} - \frac{A_{n-2}}{A_{n-1}} \cdot \frac{n-1}{n(n+1)}$$

et finalement :

$$\frac{u_n}{u_{n-1}} = \frac{n(n-1) - 4Ns}{n(n+1)} + \xi \left[\frac{1}{n+1} - \frac{n(n-1)^2}{n(n+1)[(n-2)(n-1) - 4Ns]} \right]$$

ce qui donne, toutes réductions faites :

$$\frac{u_n}{u_{n-1}} = \frac{[n(n-1) - 4Ns][(n-1)(n-2) - 4Ns] - \xi n(n-1 + 4Ns)}{n(n+1)[(n-1)(n-2) - 4Ns]}$$

ou encore :

$$(15) \quad \frac{u_{n+1}}{u_n} = \frac{[n(n+1) - 4Ns][(n-1)n - 4Ns] - \xi(n+1)(n+4Ns)}{(n+1)(n+2)[n(n-1) - 4Ns]} + O(\xi^2)$$

Le rapport u_{n+1}/u_n est de la forme

$$\frac{P(n)}{(n+1)Q(n)}$$

où $P(n)$ et $Q(n)$ sont respectivement des polynômes du quatrième et du troisième degré : la série de terme général $u_n x^n$ est donc, avec l'approximation adoptée, une *série hypergéométrique du quatrième ordre*.

Pour obtenir (toujours avec la même approximation) l'expression de u_n , cherchons des valeurs approchées des racines du numérateur et du dénominateur de (15).

Remarquons tout d'abord que $n(n-1) - 4Ns$ et $n(n+1) - 4Ns$ ont des racines respectivement opposées ; il résulte donc de (13) que le dénominateur de (15) s'écrit :

$$(n+1)(n+2)(n-\beta)(n-\gamma)$$

D'autre part les racines du numérateur de (15) sont voisines de celles de :

$$[n(n+1) - 4Ns] [n(n-1) - 4Ns]$$

c'est-à-dire de $-\beta, -\gamma, \beta, \gamma$. Désignons ces racines par b, c, b' et c' et posons par exemple

$$b = \beta + \varepsilon$$

On a, en tenant compte de :

$$\beta(\beta-1) - 4Ns = 0$$

et en négligeant $0(\varepsilon^2)$ et $0(\xi\varepsilon)$:

$$[(\beta + \varepsilon)(\beta + \varepsilon + 1) - 4Ns] [(\beta - 1 + \varepsilon)(\beta + \varepsilon) - 4Ns] - \xi(\beta + \varepsilon + 1)(\beta + \varepsilon + 4Ns) = 0$$

soit :

$$\varepsilon = \xi \cdot \frac{(\beta + 1)\beta}{2(2\beta - 1)}$$

On obtiendrait de même en posant :

$$c = \gamma + \varepsilon'$$

l'expression approchée :

$$\varepsilon' = \xi \cdot \frac{(\gamma + 1)\gamma}{2(2\gamma - 1)}$$

Posons de la même manière :

$$b' = -\beta + \varepsilon''$$

on trouve par un calcul analogue :

$$\varepsilon'' = \xi \cdot \frac{(\beta - 2)(\beta - 1)}{2(2\beta - 1)}$$

c'est-à-dire, en remarquant que $\beta + \gamma = 1$:

$$\varepsilon'' = -\varepsilon'$$

En résumé, on trouve pour les racines du numérateur de $\frac{u_{n+1}}{u_n}$ les valeurs approchées suivante

$$(16) \quad \begin{array}{l} b = \beta + \varepsilon \\ c = \gamma + \varepsilon' \\ b' = -\beta - \varepsilon' \\ c' = -\gamma - \varepsilon \end{array} \quad \text{avec :} \quad \begin{array}{l} \varepsilon = \frac{(\beta + 1)\beta}{2(2\beta - 1)} \\ \varepsilon' = \frac{(\gamma + 1)\gamma}{2(2\gamma - 1)} \end{array}$$

β et γ étant racines de : $n^2 - n - 4Ns = 0$.

On remarque que l'on a immédiatement les relations :

$$(17) \quad \boxed{\beta + \gamma = 1; \quad \beta\gamma = -4Ns; \quad \varepsilon + \varepsilon' = \xi}$$

Explicitons alors l'expression de u_n :

$$u_n = \frac{(-\beta - \varepsilon)(-\beta - \varepsilon + 1) \dots (-\beta - \varepsilon + n - 1)(-\gamma - \varepsilon')(-\gamma - \varepsilon' + 1) \dots (-\gamma - \varepsilon' + n - 1)}{(\beta + \varepsilon)(\beta + \varepsilon + 1) \dots (\beta + \varepsilon + n - 1)(\gamma + \varepsilon)(\gamma + \varepsilon + 1) \dots (\gamma + \varepsilon + n - 1)}$$

$$= \frac{1 \cdot 2 \dots n \cdot 2 \cdot 3 \dots (n+1)(-\beta)(-\beta+1) \dots (-\beta+n-1)(-\gamma)(-\gamma+1) \dots (-\gamma+n-1)}{1 \cdot 2 \dots (n-1) \cdot n \cdot 2 \cdot 3 \dots n \cdot (n+1)}$$

ce qui s'écrit :

$$u_n = \frac{\beta(\beta+1) \dots (\beta+n-1)\gamma(\gamma+1) \dots (\gamma+n-1)}{1 \cdot 2 \dots (n-1) \cdot n \cdot 2 \cdot 3 \dots n \cdot (n+1)} \cdot (1 + K\xi^2) + O(\xi^3)$$

en posant :

$$K\xi^2 = -\varepsilon \sum_{i=0}^{n-1} \left(\frac{1}{-\beta+i} \right) - \varepsilon' \sum_{i=0}^{n-1} \left(\frac{1}{-\gamma+i} \right) + \varepsilon' \sum_{i=0}^{n-1} \left(\frac{1}{\beta+i} \right) + \varepsilon \sum_{i=0}^{n-1} \left(\frac{1}{\gamma+i} \right)$$

ou, en tenant compte de (17) :

$$K\xi^2 = \varepsilon \sum_{i=0}^{n-1} [1/(\gamma+i) - 1/(\gamma+i-1)] + \varepsilon' \sum_{i=0}^{n-1} [1/(\beta+i) - 1/(\beta+i-1)]$$

soit :

$$\begin{aligned} K\xi &= \varepsilon [1/(\gamma + n - 1) - 1/(\gamma - 1)] + \varepsilon' [1/(\beta + n - 1) - 1/(\beta - 1)] \\ &= \varepsilon/\beta + \varepsilon'/\gamma + \varepsilon/(n - \beta) + \varepsilon'/(n - \gamma) \end{aligned}$$

Un calcul très simple montre à partir des relations (17) que l'on a :

$$\varepsilon/\beta + \varepsilon'/\gamma = \xi/2$$

et

$$\varepsilon/(n - \beta) + \varepsilon'/(n - \gamma) = \xi \cdot (n + 2 N_s)/(n^2 - n - 4 N_s)$$

et on obtient l'expression suivante de u_n :

$$(18) \quad u_n = A_n \left[1 + \xi \left(\frac{1}{2} + \frac{n + 2 N_s}{n^2 - n - 4 N_s} \right) \right] + O(\xi^2)$$

$A_n x^n$ étant, comme nous l'avons vu page 109, le terme général de la série hypergéométrique de Gauss de somme $\mathcal{F}(\beta, \gamma, 2, x)$.

On a donc :

$$u_n x^n = A_n x^n (1 + \xi/2) + A_n x^n \xi (n + 2 N_s)/(n^2 - n - 4 N_s)$$

Considérons la série de terme général :

$$z_n x^n = (-2/\xi) \cdot A_n x^n \xi (n + 2 N_s)/(n^2 - n - 4 N_s)$$

on a :

$$z_0 = 1$$

D'autre part :

$$\frac{z_{n+1}}{z_n} = \frac{A_{n+1}}{A_n} \cdot \frac{n + 1 + 2 N_s}{n + 2 N_s} \cdot \frac{n^2 - n - 4 N_s}{(n + 1)^2 - n - 1 - 4 N_s}$$

ou, d'après (12) :

$$\frac{z_{n+1}}{z_n} = \frac{(n + 1 + 2 N_s) (n^2 - n - 4 N_s)}{(n + 1) (n + 2) (n + 2 N_s)}$$

ce qui montre que la série de terme général $z_n x^n$ est une série hypergéométrique du troisième ordre.

Les racines du numérateur de z_{n+1}/z_n sont $\beta, \gamma, -2 N_s - 1$ et celles du dénominateur $-1, -2, -2 N_s$.

On a donc :

$$z_n = \frac{(2N_s + 1)(2N_s + 2) \dots (2N_s + n)(-\beta)(-\beta + 1) \dots (-\beta + n - 1)(-\gamma) \dots (-\gamma + n - 1)}{1 \cdot 2 \dots n \cdot 2 \cdot 3 \dots (n + 1) \cdot 2 N_s (2 N_s + 1) (2 N_s + 2) \dots (2 N_s + n - 1)}$$

ou :

$$(19) \quad z_n = \frac{2 N_s + n}{2 N_s} \cdot V_n = V_n (1 + n/2 N_s)$$

en désignant par $V_n x^n$ le terme général de la série hypergéométrique de Gauss qui a pour somme :

$$\mathcal{F}(-\beta, -\gamma, 2, x)$$

on déduit donc de (18) et (19) :

$$(20) \quad u_n x^n = A_n x^n (1 + \xi/2) - V_n x^{n\xi/2} - n V_n x^{n\xi/4} N_s$$

Remarquons alors que

$$n V_n x^n$$

n'est autre que le produit par x du terme général de la série dérivée de la série de somme $\mathcal{F}(-\beta, -\gamma, 2, x)$. Or on a :

$$\frac{d}{dx} \mathcal{F}(a, b, c, x) = \frac{ab}{c} \mathcal{F}(a + 1, b + 1, c + 1, x) \quad (1)$$

d'où, en tenant compte de (17) :

$$\frac{d}{dx} \mathcal{F}(-\beta, -\gamma, 2, x) = -2 N_s \mathcal{F}(\beta, \gamma, 3, x)$$

Désignons alors par $W_n x^n$ le terme général de la série de somme :

$$\mathcal{F}(\beta, \gamma, 3, x)$$

On a :

$$-n V_n x^{n\xi/4} N_s = -2 N_s x \cdot W_{n-1} x^{n-1} \cdot (-\xi/4 N_s) = W_{n-1} x^{n\xi/2}$$

et l'expression (20) devient finalement :

$$(21) \quad u_n x^n = A_n x^n (1 + \xi/2) - V_n x^{n\xi/2} + W_{n-1} x^{n\xi/2}$$

(1) Cf. E. GOURSAT, *Propriétés générales de l'équation d'Euler et de Gauss*, Paris, 1936, Hermann éd., p. 7.

ou :

(21')

$$u_n = A_n + (A_n - V_n + W_{n-1}) \frac{\xi}{2}$$

et l'intégrale y_1 cherchée est, avec l'approximation adoptée :

$$(22) \quad y_1 = x \mathcal{F}(\beta, \gamma, 2, x) + \frac{x\xi}{2} [\mathcal{F}(\beta, \gamma, 2, x) - \mathcal{F}(-\beta, -\gamma, 2, x) + x \mathcal{F}(\beta, \gamma, 3, x)]$$

où β et γ sont les racines de :

$$n(n-1) - 4N_s = 0$$

Remarquons que les quatre séries dont y_1 est la somme sont convergentes pour $|x| < 1$

*
* *

Proposons-nous alors de déterminer les valeurs propres s de telle sorte que y_1 s'annule pour $x = 1$.

On a pour $x = 1$:

$$y_1(1) = \sum u_n = \sum A_n (1 + \xi/2) + \xi/2 (-\sum V_n + \sum W_n)$$

Si nous supposons qu'aucune des quantités (réelles) $\beta, \gamma, 2-\beta, 2-\gamma, 3-\beta, 3-\gamma$, n'est nulle ou égale à un entier < 0 , nous avons (1) :

$$\sum A_n = \mathcal{F}(\beta, \gamma, 2, 1) = \frac{\Gamma(2) \Gamma(2-\beta-\gamma)}{\Gamma(2-\beta) \Gamma(2-\gamma)}$$

et de même :

$$\begin{cases} \sum V_n = \mathcal{F}(-\beta, -\gamma, 2, 1) = \frac{\Gamma(2) \Gamma(2+\beta+\gamma)}{\Gamma(2+\beta) \Gamma(2+\gamma)} \\ \sum W_n = \mathcal{F}(\beta, \gamma, 3, 1) = \frac{\Gamma(3) \Gamma(3-\beta-\gamma)}{\Gamma(3-\beta) \Gamma(3-\gamma)} \end{cases}$$

ce qui montre, en raison de la relation $\beta + \gamma = 1$, que :

$$\sum V_n = \sum W_n$$

(1) Cf. E. GOURSAT, *op. cit.*, p. 85 et 86.

et donc :

$$y_1(1) = (1 + \xi/2) \frac{\Gamma(2) \Gamma(2 - \beta - \gamma)}{\Gamma(2 - \beta) \Gamma(2 - \gamma)}$$

Or, par suite des égalités :

$$\beta + \gamma = 1; \quad \Gamma(p+1) = p\Gamma(p); \quad \Gamma(1) = 1$$

on a :

$$\begin{cases} \Gamma(2) = 1 \\ \Gamma(2 - \beta - \gamma) = 1 \\ \Gamma(2 - \beta) = \Gamma(1 + \gamma) = \gamma\Gamma(\gamma) \\ \Gamma(2 - \gamma) = \Gamma(1 + \beta) = \beta\Gamma(\beta) \end{cases}$$

La condition $y_1(1) = 0$ s'écrirait alors :

$$1/\Gamma(\beta) \Gamma(\gamma) = 0$$

c'est-à-dire, en raison de l'égalité : $1/\Gamma(p) \Gamma(1-p) = \sin p\pi/\pi$:

$$\frac{\sin \beta\pi}{\pi} = 0$$

ce qui exigerait que β soit un nombre entier, contrairement aux hypothèses faites sur les nombres β et γ . (En effet, si l'un des nombres β ou γ est entier, les deux le sont, et l'un des deux est ≤ 0 .)

L'hypothèse que β et γ ne sont pas entiers conduisant à une contradiction, il est nécessaire que les racines de

$$n(n-1) - 4Ns = 0$$

soient entières, et donc que s ait pour valeurs :

$$s = k(k-1)/4N$$

k étant entier. On peut d'ailleurs se borner à donner à k les valeurs entières positives, car les deux valeurs de k :

$$k = k_0 > 0 \quad \text{et} \quad k = 1 - k_0 \leq 0$$

donnent à s la même valeur. Etudions alors ce que devient y_1 :

$$1^\circ k = 1; s = 0.$$

L'équation (9) s'écrit alors :

$$x(x-1)(y'' - \xi y') = 0$$

d'où :

$$y = me^{\xi x} + p$$

la seule solution qui s'annule pour $x = 0$ et $x = 1$ est la solution triviale $y_1 = 0$.

2° $k = 2$; $s = 1/2N$.

On a alors : $\beta = 2$; $\gamma = -1$. $\mathcal{F}(2, -1, 2, x)$ et $\mathcal{F}(2, -1, 3, x)$ sont des polynômes du premier degré et $\mathcal{F}(-2, 1, 2, x)$ est un polynôme du deuxième degré (1).

On calcule facilement l'expression de y_1 à l'aide de (22) ; on trouve :

$$y_1 = x \left[1 + x \left(\frac{\xi}{2} - 1 \right) - x^2 \frac{\xi}{2} \right] + 0 (\xi^2)$$

expression qu'il est d'ailleurs aisé de retrouver directement avec l'approximation convenue en calculant les coefficients u_n à partir de la relation de récurrence (6) : on trouve effectivement que pour $n \geq 3$ on a :

$$u_n = 0 (\xi^2)$$

3° Cas général : $k = p > 2$; $s = p(p-1)/4N$.

On a alors : $\beta = p$; $\gamma = 1 - p < 0$.

$\mathcal{F}(p, 1-p, 2, x)$ et $\mathcal{F}(p, 1-p, 3, x)$ sont des polynômes de degré $p-1$; $\mathcal{F}(p-1, -p, 2, x)$ est un polynôme de degré p (1).

Il en résulte d'après (22) que y_1 est de la forme :

$$y_1 = x \cdot P(x)$$

où $P(x)$ est un polynôme de degré p : on calcule aisément le coefficient du terme de degré p de ce polynôme ; il est égal à :

$$(-1)^{p-1} \frac{(2p-2)!}{p!(p-1)!} \frac{\xi}{2} \neq 0$$

En résumé donc, les valeurs propres s_i de l'équation (9) relatives à l'intervalle $(0, 1)$ sont :

$$s_i = i(i+1)/4N \quad (i \text{ entier} > 0)$$

l'intégrale correspondant à la valeur propre s_i étant, avec l'approximation adoptée, de la forme :

$$y_1 = x P_{i+1}(x)$$

(1) Polynômes de Jacobi.

$P_{i+1}(x)$ étant un polynôme de degré $i + 1$.

Ces résultats sont évidemment conservés si l'on revient aux notations primitives en remplaçant x par x/ξ , et y_i par f_i , l'intervalle considéré étant alors $(0, \xi)$.

En particulier, les valeurs propres s_i étant positives, la fonction

$$Z(x, t) = \sum k_i e^{-s_i t} f_i(x)$$

tend vers 0 lorsque t augmente indéfiniment, comme l'exponentielle :

$$e^{-t/2 N}$$

car la plus petite valeur propre est :

$$s_1 = 1/2 N$$

Nous précisons ainsi la rapidité avec laquelle la fonction $G(x, t)$ tend vers sa limite $g(x)$.

* * *

Détermination des k_i et étude de la série $Z(x, t) = \sum_{i=1}^{\infty} k_i e^{-s_i t} f_i(x)$.

On peut écrire l'équation (9) sous forme réduite en posant :

$$h = \frac{e^{x\xi} - 1}{e^{\xi} - 1} \quad \text{d'où} \quad \frac{dh}{dx} = \frac{\xi e^{\xi x}}{e^{\xi} - 1}$$

et :

$$y(x) = T(h)$$

On obtient :

$$(23) \quad T'' + \frac{4 N s (e^{\xi} - 1)^2}{\xi^2 e^{\frac{2}{\xi} x} x (1 - x)} T = 0$$

Or on a :

$$\begin{cases} h = 0 & \text{pour } x = 0 \\ h = 1 & \text{pour } x = 1 \end{cases}$$

Nous sommes donc amenés à chercher des intégrales de (23) non nulles, et nulles pour $h = 1$ et $h = 0$.

On sait ⁽¹⁾ que la suite de telles intégrales est orthogonale relativement à la fonction :

$$H(h) = \frac{4N(e^{\frac{h}{\xi}} - 1)^2}{\xi^2 e^{\frac{h}{\xi}} x(1-x)}$$

et à l'intervalle $(0, 1)$ et est complète relativement à cet intervalle. C'est-à-dire que l'on a :

$$(24) \quad \boxed{\int_0^1 HT_i T_j dh = 0 \quad i \neq j}$$

et que, d'autre part, si $Y(t)$ est une fonction continue dans $(0, 1)$ la condition :

$$\int_0^1 YT_i dt = 0$$

entraîne, si elle est réalisée quel que soit i , la nullité de $Y(t)$ en tout point de $(0, 1)$.

Ces remarques permettent de calculer les nombres k_i . En effet, par suite de la condition initiale :

$$G(x, 0) = e^{p_0 x}$$

on doit avoir :

$$(25) \quad \boxed{Z(x, 0) = G(x, 0) - g(x) = \sum_{i=1}^{\infty} k_i f_i(x) = e^{p_0 x} - \frac{e^{p_0 \xi} - 1}{e^{\xi} - 1} (e^x - 1) - 1}$$

ou, en tenant compte des changements de variables successifs que nous avons faits $[x/\xi \text{ puis } h(x/\xi)]$:

$$(25'') \quad \boxed{\sum_{i=1}^{\infty} k_i T_i(h) = e^{p_0 \xi x} - \frac{e^{p_0 \xi} - 1}{e^{\xi} - 1} (e^{\xi x} - 1) - 1}$$

(1) Cf. E. PICARD, *Quelques problèmes aux limites de la théorie des équations différentielles*, Paris, Gauthier-Villars éd., p. 36 et 67.

Si l'on pose :

(26)

$$\int_0^1 H(h) T_i^2 dh = M_i$$

on obtient à partir de (25'), en tenant compte de (24) et de (26) :

$$k_i = \int_0^1 \frac{Z(h, 0) T_i(h) H(h) dh}{M_i}$$

où $Z(h, 0)$ représente l'expression (25'). Et, en revenant à la variable x qui figure dans (9) et (22) :

(27)

$$k_i = \int_0^1 \frac{4 N(e^{\xi} - 1)}{\xi e^{\xi x} x (1 - x)} \left[e^{p_0 \xi x} - \frac{e^{p_0 \xi} - 1}{e^{\xi} - 1} (e^{\xi x} - 1) - 1 \right] \frac{y_i}{M_i} dx$$

M_i étant donné par :

(26')

$$M_i = \int_0^1 \frac{4 N(e^{\xi} - 1)}{\xi e^{\xi x} x (1 - x)} y_i^2 dx$$

Les intégrales (27) et (26') ont toujours un sens, les polynômes y_i étant tous divisibles par $x(1 - x)$.

Les quantités k_i étant ainsi déterminées, on sait ⁽¹⁾ que la série $\sum |k_i T_i(h)|$ est uniformément convergente dans l'intervalle $(0, 1)$, que la série $\sum k_i T_i$ a pour somme $Z(h, 0)$ et enfin que la série $\sum k_i T_i e^{-s_i t}$, uniformément convergente par rapport à h ($0 < h < 1$) et t ($t \geq 0$) est dérivable par rapport à t et admet une dérivée seconde par rapport à h .

En revenant à la variable x on a donc :

(28)

$$\frac{\partial Z(x, t)}{\partial t} = \sum -s_i k_i e_i - s_i t f_i(x)$$

(1) Cf. E. PICARD, *op. cit.*, p. 79 et suivantes.

Or, on a d'après (7) :

$$f_i' - f_i'' = -4 N_{si} f_i / x (x - \xi)$$

et en dérivant :

$$f_i'' - f_i''' = -4 N_{si} f_i' / x (x - \xi) + 4 N_{si} f_i (2x - \xi) / x^2 (x - \xi)^2$$

L'intégrale (33) s'écrit donc :

$$(33') \quad S(x, t) = \frac{4N}{x(x-\xi)} \left[\sum k_{ie}^{-s_i t} \left[f_i + 2f_i' - 2f_i \frac{(2x-\xi)}{x(x-\xi)} \right] \right]_0^t$$

On a (cf. p. 118) :

$$\sum k_{if_i} = e^{p_{0x}} - \frac{e^{p_{0\xi}} - 1}{e^{\xi} - 1} (e^x - 1) - 1$$

et en dérivant :

$$\sum k_{if_i'} = p_{0e} p_{0x} - \frac{e^{p_{0\xi}} - 1}{e^{\xi} - 1} e^x$$

Donc pour $t = 0$ l'expression entre crochets du deuxième membre de (33') est égale à :

$$(34) \quad R(x) = (1 + 2p_0) e^{p_{0x}} - 3a'e^x + a' - 1 - 2 \left[e^{p_{0x}} - a'(e^x - 1) - 1 \right] \frac{2x - \xi}{x(x - \xi)}$$

en ayant posé :

$$(35) \quad a' = \frac{e^{p_{0\xi}} - 1}{e^{\xi} - 1}$$

et l'équation (32) s'écrit :

$$(36) \quad F(x, t) = e^{p_{0x}} + \frac{4N\alpha^2\mu}{x-\xi} R(x) + \int_0^t \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) (x - \xi) x / 4N dt - \frac{4N\alpha^2\mu}{x-\xi} \sum k_{ie}^{-s_i t} \left[f_i + 2f_i' - 2f_i \frac{(2x-\xi)}{x(x-\xi)} \right]$$

Proposons-nous de trouver la limite $\varphi(x)$ de $F(x, t)$ lorsque t augmente indéfiniment. Nous avons vu (p. 103) qu'elle est de la forme

$$\varphi(x) = a(e^x - 1) + 1$$

et d'autre part pour t augmentant indéfiniment, (36) devient :

$$\varphi(x) = e^{p_0 x} + \frac{4 N \alpha^2 \mu}{x - \xi} R(x) + \int_0^\infty \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) (x - \xi) x / 4N dt$$

Pour déterminer $\varphi(x)$, c'est-à-dire le coefficient a , il suffit de connaître la valeur de $\varphi(x)$ pour une valeur particulière de x différente de 0. Or, si l'on remarque que pour $x = \xi$ l'intégrale qui figure dans l'expression précédente de $\varphi(x)$ est nulle, on peut écrire, à condition que cette intégrale soit continue pour $x = \xi$:

$$(37) \quad \varphi(\xi) = e^{p_0 \xi} + \lim_{x \rightarrow \xi} \frac{4 N \alpha^2 \mu R(x)}{x - \xi}$$

Montrons donc la continuité pour $x = \xi$ de l'intégrale

$$I(x) = \int_0^\infty \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) (x - \xi) x / 4N dt$$

Supposons par exemple $\xi > 0$. Dans l'intervalle $(0, \xi)$, la quantité $(x - \xi)$ est < 0 .

Les inégalités (17) établies au chapitre I (p. 96), s'écrivent, si l'on tient compte de l'expression (2) de ξ :

$$\left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) (x - \xi - 12 N \alpha^2 \mu) x / 4N \leq \frac{\partial F}{\partial t} \leq \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) (x - \xi - 4 N \alpha^2 \mu) x / 4N$$

Les quantités $x - \xi - 12 N \alpha^2 \mu$ et $x - \xi - 4 N \alpha^2 \mu$ sont < 0 , quelles que soient les hypothèses faites sur ξ , lorsque x est suffisamment proche de ξ . Donc :

$$\frac{\partial F}{\partial t} \cdot \frac{4 N}{x(x - \xi - 12 N \alpha^2 \mu)} \leq \frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \leq \frac{\partial F}{\partial t} \cdot \frac{4 N}{x(x - \xi - 4 N \alpha^2 \mu)}$$

ou enfin, puisque $x(x - \xi)$ est < 0 :

$$\frac{\partial F}{\partial t} \cdot \frac{x - \xi}{x - \xi - 4 N \alpha^2 \mu} \leq \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) x(x - \xi) / 4N \leq \frac{\partial F}{\partial t} \cdot \frac{x - \xi}{x - \xi - 12 N \alpha^2 \mu}$$

et donc :

$$\int_0^{\infty} \frac{x - \xi}{x - \xi - 4 N \alpha^2 \mu} \cdot \frac{\partial F}{\partial t} dt \leq I(x) \leq \int_0^{\infty} \frac{x - \xi}{x - \xi - 12 N \alpha^2 \mu} \cdot \frac{\partial F}{\partial t} dt$$

ou encore :

$$[\varphi(x) - F(x, 0)] \cdot \frac{x - \xi}{x - \xi - 4 N \alpha^2 \mu} \leq I(x) \leq [\varphi(x) - F(x, 0)] \cdot \frac{x - \xi}{x - \xi - 12 N \alpha^2 \mu}$$

Les deux membres extrêmes étant des fonctions continues de x qui tendent vers 0 lorsque x tend vers ξ , $I(x)$ tend bien vers 0 lorsque x tend vers ξ .

Le raisonnement précédent s'applique évidemment encore au cas où ξ est < 0 ; il suffit d'inverser le sens des inégalités.

$\varphi(\xi)$ a donc bien l'expression limite (37). Pour la calculer, posons :

$$x = \xi + \varepsilon.$$

d'où :

$$4 N \alpha^2 \mu / \varepsilon R(\xi + \varepsilon) = 4 N \alpha^2 \mu / \varepsilon \left\{ -\frac{2(\xi + 2\varepsilon)}{\varepsilon(\xi + \varepsilon)} [e^{p_0(\xi + \varepsilon)} - a', (e^{\xi + \varepsilon} - 1) - 1] \right. \\ \left. + (1 + 2p_0)e^{p_0(\xi + \varepsilon)} - 3a'e^{\xi + \varepsilon} + a' - 1 \right\}$$

ou :

$$\frac{4 N \alpha^2 \mu}{\varepsilon} R(\xi + \varepsilon) = 4 N \alpha^2 \mu / \varepsilon \left\{ -2(1/\varepsilon + 1/\xi - \varepsilon/\xi^2) e^{p_0\xi} (1 + p_0\varepsilon + p_0^2\varepsilon^2/2) \right. \\ \left. + 2(1/\varepsilon + 1/\xi - \varepsilon/\xi^2) a'e^{\xi} (1 + \varepsilon + \varepsilon^2/2) - a' + 1 \right\} \\ + (1 + 2p_0)e^{p_0\xi} (1 + p_0\varepsilon) - 3a'e^{\xi} (1 + \varepsilon) + a' - 1 \\ + 0(\varepsilon)$$

ce qui donne toutes réductions faites :

$$\frac{4 N \alpha^2 \mu}{\varepsilon} R(\xi + \varepsilon) = 4 N \alpha^2 \mu [2a'e^{\xi} (1/\xi - 1) + p_0e^{p_0\xi} (1 + p_0 - 2/\xi)] + 0(\varepsilon)$$

et finalement :

$$(38) \quad \varphi(\xi) = e^{p_0\xi} + 4 N \alpha^2 \mu [2a'e^{\xi} (1/\xi - 1) + p_0e^{p_0\xi} (1 + p_0 - 2/\xi)]$$

Or, on a d'après (35) :

$$a' = p_0 + (p_0 - 1) p_0 \xi / 2 + p_0 (p_0 - 1) (2 p_0 - 1) \xi^2 / 12 + 0(\xi^3)$$

d'où tous calculs faits :

$$\varphi(\xi) = 1 + p_0 \xi + p_0^2 \xi^2 / 2 + p_0^3 \xi^3 / 6 + 4 N \alpha^2 \mu p_0 (p_0 - 1) (2 p_0 + 5) / 6 + 0 (\alpha^4)$$

égalité qui permet de calculer a . On a en effet :

$$\varphi(\xi) = a (e^\xi - 1) + 1 = a \xi (1 + \xi/2 + \xi^2/6) + 1 + 0 (\xi^4)$$

et donc :

$$a (1 + \xi/2 + \xi^2/6) = p_0 + p_0^2 \xi/2 + p_0^3 \xi^2/6 + 4 N \alpha^2 \mu p_0 (p_0 - 1) (2 p_0 + 5) / 6 + 0 (\alpha^3)$$

ou :

$$a = p_0 [1 - (1 - p_0) \xi/2 + 4 N \alpha^2 \mu (p_0 - 1) (2 p_0 + 5) / 6 + \xi^2 (2 p_0^2 - 3 p_0 + 1) / 12] + 0 (\alpha^3)$$

Si on suppose que le produit μN est petit de telle sorte que l'on puisse négliger $0 (\xi^2)$ par rapport à $0 (N \mu \alpha^2)$, on obtient finalement en remplaçant ξ par son expression (2), la valeur suivante pour a :

$$(39) \quad a = p_0 + 2 N \alpha \mu p_0 (1 - p_0) [2 \underline{0} - (2 p_0 + 5) \alpha / 3] + 0 (\alpha^3) + 0 (N^2 \mu^2 \alpha^2)$$

Remarquons que si l'on pose :

$$(40) \quad \eta = - 4 N \alpha \mu [2 \underline{0} - \alpha \frac{2 p_0 + 5}{3}]$$

et si l'on tient compte du développement :

$$\frac{e^{p_0 \eta} - 1}{e^\eta - 1} = p_0 - p_0 (1 - p_0) \eta / 2 + 0 (\eta^2)$$

l'expression (39) donne, avec l'approximation adoptée :

$$(41) \quad a = \frac{e^{p_0 \eta} - 1}{e^\eta - 1}$$

Ceci montre que la fonction $F(x, t)$ a même limite $\varphi(x)$ lorsque t augmente indéfiniment que la solution de l'équation :

$$(42) \quad \frac{\partial F}{\partial t} = \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) x (x - \eta)$$

qui satisfait à la même condition initiale.

Cette équation (42) est analogue à l'équation (3) et sa solution approchée s'obtiendrait comme nous l'avons vu.

*
* *

Comparaison avec le random mating :

M. MALÉCOT a montré ⁽¹⁾ que dans une population limitée, la panmixie aboutit nécessairement à l'homogénéité. Nous venons de voir qu'il en est de même si la population est soumise à notre loi d'union sélective. Mais, alors que dans le cas du random mating la probabilité de fixation du gène A est égale à p_0 , elle est ici égale à a .

L'expression de a montre immédiatement que l'on a :

$$a > p_0$$

si η est < 0 . Or η a le signe de $-\underline{Q}$ puisque par hypothèse $\underline{Q}$ est (en valeur absolue) grand par rapport à α .

Donc l'union sélective favorise la fixation du gène A si l'optimum $\underline{Q}$ est > 0 , c'est-à-dire si l'homozygote AA est le moins mal adapté des trois génotypes. Elle favorise au contraire la fixation du gène récessif a si l'homozygote aa est le moins mal adapté ($\underline{Q} < 0$).

(1) Cf. G. MALÉCOT, *Les mathématiques de l'hérédité*, p. 30, Paris, 1948, Masson éd.

CHAPITRE III

EVOLUTION D'UNE POPULATION LIMITÉE RELATIVEMENT BIEN ADAPTÉE

Si la population est bien adaptée, l'optimum $\underline{O}$ est de l'ordre de grandeur de α , contribution du gène A au caractère primaire ; $\underline{O}$ est donc petit.

L'équation (E) s'écrit alors :

$$(1) \quad \frac{\partial F}{\partial t} = \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) x^2 / 4N + 2 \underline{O} \alpha \mu x \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) - \alpha^2 \mu x \left[\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} + 2 \left(\frac{\partial^2 F}{\partial x^2} - \frac{\partial^3 F}{\partial x^3} \right) \right]$$

Le deuxième et le troisième termes du deuxième membre sont petits par rapport au premier. Nous prendrons comme première approximation de la fonction $F(x, t)$ la fonction $G(x, t)$ définie par :

$$(2) \quad \frac{\partial G}{\partial t} = \left(\frac{\partial G}{\partial x} - \frac{\partial^2 G}{\partial x^2} \right) x^2 / 4N$$

satisfaisant à la condition initiale :

$$(3) \quad G(x, 0) = e^{p_0 x} \quad \text{d'où (3')} \quad \frac{\partial G(x, 0)}{\partial x} = p_0 e^{p_0 x}$$

La forme de l'équation (2) montre que pour $x = 0$ on a :

$$\frac{\partial G}{\partial t} = 0 \quad \text{et} \quad \frac{\partial^2 G}{\partial x \partial t} = 0$$

Il en résulte que l'on a :

$$(4) \quad \begin{aligned} G(0, t) &= G(0, 0) = 1 \\ \frac{\partial G(0, t)}{\partial x} &= \frac{\partial G(0, 0)}{\partial x} = p_0 \end{aligned}$$

L'équation (2) admet des solutions stationnaires $S(x)$ données par :

$$\frac{dS}{dx} - \frac{d^2S}{dx^2} = 0$$

soit :

$$S(x) = ae^x + b \quad \text{et} \quad \frac{dS}{dx} = ae^x$$

En particulier la solution stationnaire qui satisfait aux conditions (4) est :

$$(5) \quad g(x) = p_0 e^x + 1 - p_0$$

Posons alors :

$$G(x, t) - g(x) = Z(x, t)$$

$Z(x, t)$ est une solution de (2) qui est nulle, ainsi que $\frac{\partial Z}{\partial x}$ pour $x = 0$. Soit :

$$Z(x, t) = \sum k_i f_i(x) h_i(t)$$

chaque produit $f_i h_i$ vérifiant (2) et f_i s'annulant ainsi que sa dérivée pour $x = 0$.

f et h doivent alors satisfaire à l'équation :

$$h'/h = (f'/f - f''/f) x^2/4N = -s$$

d'où finalement :

$$(6) \quad G(x, t) = p_0 (e^x - 1) + 1 + \sum k_i e^{-s_i t} f_i(x)$$

f_i étant solution de l'équation différentielle :

$$(7) \quad x^2 (f_i'' - f_i') - 4N s_i f_i = 0$$

les constantes k_i et s_i devant être telles que la condition initiale (3) soit satisfaite ainsi que les conditions :

$$(8) \quad f_i(0) = 0 \quad \text{et} \quad (8') \quad f_i'(0) = 0$$

* * *

Remarque. — L'équation (2) est celle que l'on obtiendrait dans le cas du random mating ($\mu = 0$). La fonction $G(x, t)$ est donc la fonction caractéristique de la fréquence p du gène A dans une population panmictique sans sélection.

Détermination des fonction $f_i(x)$:

L'équation (7) est du type de Fuchs. En posant :

$$f = x^r + u_1 x^{r+1} + \dots + u_n x^{r+n} + \dots$$

on obtient, après substitution dans (7), en annulant le coefficient de x^r , l'équation déterminante :

(9)

$$r(r-1) - 4 N_{s_i} = 0$$

C'est précisément l'équation dont les racines sont les nombres β et γ qui sont intervenus au chapitre précédent (cf. équation (13) p. 109).

Or, le cas que nous étudions ici doit être considéré comme le cas limite (pour $\xi = 0$) du cas considéré au chapitre II : il ne peut en être ainsi que si les solutions trouvées alors ne cessent pas d'exister, c'est-à-dire si la constante s_i qui figure dans l'équation (7) est égale à une valeur propre de l'équation (II-7, p. 106) du chapitre précédent, ces valeurs propres étant, comme nous l'avons vu, indépendantes de ξ .

On doit donc avoir ici également :

$$s_i = i(i+1)/4N \quad (i \geq 1)$$

et l'équation (9) a les racines entières :

$$\beta = i+1 \quad \gamma = -i$$

Comme la différence de ces racines est entière ($\beta - \gamma = 2i+1$) l'équation (7) admet les deux solutions :

$$y_1 = x^{i+1} \psi(x)$$

$$y_2 = x^{i+1} \psi(x) [C \log x + x^{-(2i+1)} \chi(x)]$$

avec : $\psi(0) \neq 0$ et $\chi(0) \neq 0$.

La solution générale :

$$C_1 y_1 + C_2 y_2$$

ne peut être nulle pour $x = 0$ que si l'on a : $C_2 = 0$. Si l'on remarque que la dérivée :

$$y_i' = x^{i+1} \psi'(x) + (i+1) x^i \psi(x)$$

est nulle pour $x = 0$, les conditions (8) et (8') sont satisfaites pour y_1 et on a :

$$(10) \quad f_i = x^{i+1} (1 + u_{i1}x + u_{i2}x^2 + \dots + u_{i,n}x^n + \dots)$$

d'où :

$$f_i = x^{i+1} + u_{i1}x^{i+2} + \dots + u_{i,n-1}x^{i+n} + \dots$$

$$f_i' = (i+1)x^i + \dots + (i+n-1)u_{i,n-2}x^{i+n-2} + \dots$$

$$f_i'' = i(i+1)x^{i-1} + \dots + (i+n-1)(i+n)u_{i,n-1}x^{i+n-2} + \dots$$

et, en annulant le coefficient de x^{i+n} après substitution des expressions précédentes dans (7), on obtient la récurrence :

$$u_{i,n-1} [(i+n-1)(i+n) - 4Ns_i] = u_{i,n-2}(i+n-1)$$

ou encore :

$$\frac{u_{i,n+1}}{u_{i,n}} = \frac{i+n+1}{(i+n+1)(i+n+2) - 4Ns_i}$$

ce qui s'écrit finalement en tenant compte des valeurs de s_i et de β :

$$(11) \quad \frac{u_{i,n+1}}{u_{i,n}} = \frac{n+i+1}{(n+1)(n+2i+2)} = \frac{n+\beta}{(n+1)(n+2\beta)}$$

Ce qui montre que la fonction $f_i(x)$ est égale à :

$$(12) \quad f_i(x) = x^{i+1} \mathcal{G}(i+1, 2i+2, x)$$

$\mathcal{G}(a, b, x)$ représentant la somme de la série de Kummer de terme général :

$$\frac{a(a+1)(a+2) \dots (a+n-1)}{1 \cdot 2 \cdot 3 \dots (n-1) n \cdot b(b+1)(b+2) \dots (b+n-1)} \cdot x^n$$

La fonction $G(x, t)$ s'écrit donc :

$$(13) \quad G(x, t) = p_0 e^x + 1 - p_0 + \sum k_i e^{-i(i+1)t/4N} x^{i+1} \mathcal{G}(i+1, 2i+2, x)$$

expression qui montre que si t augmente indéfiniment, $G(x, t)$ a pour limite $g(x)$; on retrouve ainsi le fait que, s'il n'y a pas de sélection (cas du random mating), la population tend vers l'homogénéité, la probabilité de survivance du gène A étant p_0 ; la vitesse de disparition de l'un des gènes étant caractérisée par la rapidité de décroissance de la fonction exponentielle :

$$e^{-t/2N}$$

* * *

Calcul des k_i . — Etude de la série :

$$\sum k_i e^{-i(i+1)t/4N} \mathcal{G}(i+1, 2i+2, x)$$

La condition initiale (3) donne :

$$Z(x, 0) = e^{p_0 x} - p_0 e^x = 1 + p_0$$

Nous nous proposons de calculer les quantités k_i de telle sorte que la série considérée soit convergente et admette pour somme $Z(x, 0)$ pour $t = 0$.

En développant $Z(x, 0)$ en série entière, on obtient :

$$Z(x, 0) = (p_0^2 - p_0) x^2/2! + (p_0^3 - p_0) x^3/3! + \dots + (p_0^n - p_0) x^n/n! + \dots$$

Considérons la série de terme général :

$$k_i x^{i+1} \mathcal{G}(i+1, 2i+2, x) = U_i$$

En désignant toujours par $u_{i,n}$ le terme général de la série de Kummer $\mathcal{G}(i+1, 2i+2, x)$, la somme des coefficients des termes de degré n (en nombre fini) dans la série $\sum U_i$ est :

$$\sum_{i=1}^{n-1} k_i u_{i, n-1-i}$$

Nous définirons les k_i par la condition que cette somme soit égale au coefficient de x^n dans le développement de $Z(x, 0)$. On a donc :

$$(14) \quad (p_0^n - p_0)/n! = \sum_{i=1}^{n-1} k_i u_{i, n-1-i}$$

Ces relations permettent de calculer de proche en proche les quantités k_i . On a en effet :

$$(14') \quad k_n = \frac{p_0^{n+1} - p_0}{(n+1)!} - k_1 u_{1, n-1} - k_2 u_{2, n-2} - \dots - k_{n-1} u_{n-1, 1}$$

puisque, quel que soit i on a : $u_{i, 0} = 1$.

Le calcul donne immédiatement :

$$\begin{cases} k_1 = -p_0(1 - p_0)/2 \\ k_2 = p_0(1 - p_0)(1 - 2p_0)/12 \\ k_3 = p_0(1 - p_0)(-5p_0^2 + 5p_0 - 1)/120 \\ \text{etc...} \end{cases}$$

Or, remarquons que si j est différent de 0 on a :

$$0 < u_{i,j} = \frac{(i+1)(i+2) \dots (i+j)}{(2i+2)(2i+3) \dots (2i+j+1) \cdot j!} \leq \frac{1}{2} \cdot \frac{1}{j!}$$

Comme d'autre part on a quel que soit p_0 :

$$|p_0^{n+1} - p_0| < 1$$

on déduit de (14') l'inégalité :

$$(15) \quad |k_n| < \frac{1}{n+1!} + |k_1| \cdot \frac{1}{2} \cdot \frac{1}{n-1!} + |k_2| \cdot \frac{1}{2} \cdot \frac{1}{n-2!} + \dots + |k_{n-1}| \cdot \frac{1}{2} \cdot \frac{1}{1!}$$

En faisant la somme membre à membre des n premières inégalités (15) on trouve :

$$\sum_{i=1}^n |k_i| < \sum_{i=1}^n 1/i+1! + \frac{1}{2} \left[|k_1| \sum_{i=1}^{n-1} 1/i! + |k_2| \sum_{i=1}^{n-2} 1/i! + \dots + |k_{n-1}| \right]$$

et donc *a fortiori* :

$$\sum_{i=1}^n |k_i| < e - 2 + (e - 1)/2 \sum_{i=1}^n |k_i|$$

et finalement :

$$\sum_{i=1}^n |k_i| < 2(e - 2)/(3 - e)$$

ce qui montre que la série $\sum_{i=1}^{\infty} k_i$ est absolument convergente.

Or on a :

$$|\mathcal{G}(i+1, 2i+2, x)| \leq 1 + \sum_{j=1}^{\infty} u_{i,j} |x|^j < \frac{1}{2} + \frac{1}{2} \sum_{j=0}^{\infty} |x|^j/j!$$

ou :

$$|\mathcal{G}(i+1, 2i+2, x)| < (1 + e^{|x|})/2$$

On peut écrire par conséquent :

$$\begin{aligned} \sum_{i=1}^{\infty} |U_i| &= \sum_{i=1}^{\infty} |k_i f_i| \leq \sum_{i=1}^{\infty} |k_i| |x|^{i+1} \mathcal{G}(i+1, 2i+2, |x|) \\ &< (1 + e^{|x|})/2 \cdot \sum_{i=1}^{\infty} |k_i| |x|^{i+1} \end{aligned}$$

Cette dernière série est convergente au moins pour $|x| \leq 1$. Il en résulte que la série de terme général U_i est absolument convergente au moins dans l'intervalle $|x| \leq 1$.

Remarquons d'ailleurs que la série est uniformément convergente dans cet intervalle puisque la série :

$$(1 + e^{|x|})/2 \cdot \sum_{i=1}^{\infty} |k_i| |x|^{i+1}$$

l'est.

La convergence de la série :

$$\sum_{i=1}^{\infty} |k_i| |x|^{i+1} \mathcal{S}(i+1, 2i+2, |x|)$$

montre que la série double :

$$\sum_{i=1}^{\infty} k_{ix}^{i+1} \sum_{j=0}^{\infty} u_{i,j} x^j$$

est absolument convergente ; on peut donc, sans changer sa somme, grouper ses termes dans un ordre arbitraire ; en particulier, en posant :

$$i + j + 1 = n$$

on peut la mettre sous la forme :

$$\sum_{n=2}^{\infty} \lambda_n x^n$$

où λ_n d'après (14) n'est autre que le coefficient de x^n dans le développement de $Z(x, 0)$.

La série ΣU_i a donc bien pour somme $Z(x, 0)$, au moins dans l'intervalle $|x| \leq 1$.

On peut voir directement que la série ΣU_i peut être dérivée terme à terme. Posons en effet :

$$\frac{dU_i}{dx} = V_i + W_i$$

avec :

$$V_i = (i+1) k_i x^i \mathcal{S}(i+1, 2i+2, x)$$

$$W_i = k_{ix}^{i+1} \frac{d \mathcal{S}(i+1, 2i+2, x)}{dx}$$

Il suffit de montrer que les deux séries ΣV_i et ΣW_i sont uniformément convergentes.

Or, d'après ce qui précède, on a :

$$\Sigma |V_i| \leq \Sigma |k_i| (i+1) |x|^i \mathcal{S}(i+1, 2i+2, |x|) < (1 + e^{|x|})/2 \cdot \Sigma (i+1) |k_i| |x|^i$$

La série ΣV_i est donc uniformément convergente au moins pour $|x| < 1$, puisque

$$(1 + e^{|x|})/2 \cdot \Sigma (i+1) |k_i| |x|^i$$

l'est : $(i+1) k_{ix}^i$ est en effet le terme général de la série dérivée de la série entière Σk_{ix}^{i+1} convergente pour $|x| \leq 1$.

Considérons de même ΣW_i . On a :

$$\frac{d \mathcal{L}(i+1, 2i+2, x)}{dx} = \frac{i+1}{2i+2} \mathcal{L}(i+2, 2i+3, x) = \frac{\mathcal{L}(i+2, 2i+3, x)}{2}$$

Or, en posant :

$$\mathcal{L}(i+1, 2i+2, x) = \Sigma u_{i,j} x^j \quad \text{et} \quad \mathcal{L}(i+2, 2i+3, x) = \Sigma w_{i,j} x^j$$

on a immédiatement :

$$\frac{w_{i,j}}{u_{i,j}} = \frac{(2i+2)(i+j+1)}{(2i+j+2)(i+1)} = \frac{2i+2j+2}{2i+j+2} < 2$$

On peut donc écrire :

$$2 \Sigma |W_i| \leq \Sigma |k_i| |x|^{i+1} \mathcal{L}(i+2, 2i+3, |x|) < 2 \Sigma |k_i| |x|^{i+1} \mathcal{L}(i+1, 2i+2, |x|)$$

cette dernière série étant uniformément convergente, la série ΣW_i l'est également.

On verrait de manière analogue que la série $\sum \frac{dU_i}{dx}$ est dérivable terme à terme, et ainsi de suite...

La série de terme général :

$$|k_i| e^{-i(i+1)t/4N} |x|^{i+1} |\mathcal{L}(i+1, 2i+2, x)|$$

est majorée par la série $\Sigma |U_i|$. Il en résulte que la série :

$$Z(x, t) = \sum k_i e^{-i(i+1)t/4N} x^{i+1} \mathcal{L}(i+1, 2i+2, x)$$

est également convergente uniformément par rapport à x et dérivable par rapport à x au moins pour $|x| < 1$. On a par conséquent :

$$(16) \quad \begin{aligned} \frac{\partial G(x, t)}{\partial x} &= p_0 e^x + \sum_{i=1}^{\infty} k_i e^{-i(i+1)t/4N} N_{f_i}'(x) \\ \frac{\partial^2 G(x, t)}{\partial x^2} &= p_0 e^x + \sum_{i=1}^{\infty} k_i e^{-i(i+1)t/4N} N_{f_i}''(x) \\ \frac{\partial^3 G(x, t)}{\partial x^3} &= p_0 e^x + \sum_{i=1}^{\infty} k_i e^{-i(i+1)t/4N} N_{f_i}'''(x) \end{aligned}$$

Il nous reste enfin à montrer que la série $Z(x, t)$, qui est évidemment uniformément convergente par rapport à t ($t \geq 0$), est dérivable terme à terme par rapport à t .

Il suffit pour cela de démontrer la convergence uniforme (pour $|x| < 1$) de la série de terme général :

$$-i(i+1)/4N k_i x^{i+1} \mathcal{G}(i+1, 2i+2, x)$$

Or, on a d'après ce qui précède :

$$\begin{aligned} \sum \left| \frac{-i(i+1)}{4N} k_i x^{i+1} \mathcal{G}(i+1, 2i+2, x) \right| &< \sum \frac{i(i+1)}{4N} |k_i| |x|^{i+1} |\mathcal{G}(i+1, 2i+2, x)| \\ &< \frac{x^2}{4N} \cdot \frac{1+e^{|x|}}{2} \cdot \sum i(i+1) |k_i| |x|^{i-1} \end{aligned}$$

Or, cette dernière série est uniformément convergente au moins pour $|x| < 1$, puisque $i(i+1) k_i x^{i-1}$ est le terme général de la dérivée seconde de la série entière $\sum k_i x^{i+1}$ convergente au moins pour $|x| \leq 1$.

On a donc finalement :

$$(16') \quad \frac{\partial G(x, t)}{\partial t} = \sum_{i=1}^{\infty} -i(i+1) k_i e^{-i(i+1)t/4N} f_i(x)/4N$$

*
* *
*

Pour obtenir une meilleure approximation de la fonction $F(x, t)$, nous tiendrons compte des termes jusqu'alors négligés de l'équation (1) en substituant à la fonction F , dans ces termes correctifs, la fonction $G(x, t)$ que nous venons de déterminer.

On obtient :

$$(17) \quad \frac{\partial F}{\partial t} = \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) x^2/4N + 2Q_{\alpha\mu x} \left(\frac{\partial G}{\partial x} - \frac{\partial^2 G}{\partial x^2} \right) - \alpha^2 \mu x \left[\frac{\partial G}{\partial x} - \frac{\partial^2 G}{\partial x^2} + 2 \left(\frac{\partial^2 G}{\partial x^2} - \frac{\partial^3 G}{\partial x^3} \right) \right]$$

d'où en tenant compte de (16) et (16') et de la condition initiale :

$$\begin{aligned}
 F(x, t) = e^{p_0 x} + \int_0^t \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) \frac{x^2}{4N} dt \\
 + (2 \underline{Q}_{\alpha\mu} - \alpha^2 \mu) x \int_0^t \sum k_{ie} e^{-s_i t} (f_i' - f_i'') dt \\
 + 2 \alpha^2 \mu x \int_0^t \sum k_{ie} e^{-s_i t} (f_i''' - f_i'') dt
 \end{aligned}$$

et en dérivant par rapport à x :

$$\begin{aligned}
 \frac{\partial F}{\partial x} = p_0 e^{p_0 x} + \int_0^t \left(\frac{\partial^2 F}{\partial x^2} - \frac{\partial^3 F}{\partial x^3} \right) \frac{x^2}{4N} dt + \int_0^t \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) \frac{x}{2N} dt \\
 + (2 \underline{Q}_{\alpha\mu} - \alpha^2 \mu) \int_0^t \sum k_{ie} e^{-s_i t} (f_i' - f_i'') dt \\
 + [(2 \underline{Q}_{\alpha\mu} - \alpha^2 \mu) x - 2 \alpha^2 \mu] \int_0^t \sum k_{ie} e^{-s_i t} (f_i'' - f_i''') dt \\
 - 2 \alpha^2 \mu x \int_0^t \sum k_{ie} e^{-s_i t} (f_i''' - f_i''') dt
 \end{aligned}$$

On a donc pour $x = 0$:

$$\begin{aligned}
 \frac{\partial F(0, t)}{\partial x} = p_0 + (2 \underline{Q}_{\alpha\mu} - \alpha^2 \mu) \int_0^t \sum k_{ie} e^{-s_i t} [f'(0) - f''(0)] dt \\
 - 2 \alpha^2 \mu \int_0^t \sum k_{ie} e^{-s_i t} [f''(0) - f'''(0)] dt
 \end{aligned}$$

ou encore :

$$(18) \quad \frac{\partial F(0, t)}{\partial x} = p_0 + I(0, t)$$

en posant :

$$(19) \quad I(x, t) = (2 \underline{O}^{\alpha\mu} - \alpha^2 \mu) \int_0^t \sum k_{ie}^{-s_i t} (f_i' - f_i'') dt - 2 \alpha^2 \mu \int_0^t \sum k_{ie}^{-s_i t} (f_i'' - f_i''') dt$$

On a :

$$\int_0^t \sum k_{ie}^{-s_i t} (f_i' - f_i'') dt = - \sum \frac{k_{ie}^{-s_i t} (f_i' - f_i'')}{s_i} + \sum \frac{k_{ie} (f_i' - f_i'')}{s_i}$$

et :

$$\int_0^t \sum k_{ie}^{-s_i t} (f_i'' - f_i''') dt = - \sum \frac{k_{ie}^{-s_i t} (f_i'' - f_i''')}{s_i} + \sum \frac{k_{ie} (f_i'' - f_i''')}{s_i}$$

Or d'après (3) et (6) :

$$(20) \quad \sum k_{ie} f_i = e^{p_0 x} - p_0 e^x + p_0 - 1$$

et on a en dérivant :

$$(21) \quad \sum k_{ie} f_i' = p_0 e^{p_0 x} - p_0 e^x$$

D'autre part, on déduit de l'équation (7) :

$$(f_i'' - f_i')/s_i = 4 N f_i / x^2$$

et en dérivant :

$$(f_i''' - f_i'')/s_i = 4 N f_i' / x^2 - 2 N f_i / x^3$$

et donc d'après (20) et (21) :

$$\sum \frac{k_{ie} (f_i' - f_i'')}{s_i} = - (e^{p_0 x} - p_0 e^x + p_0 - 1) 4 N / x^2$$

et :

$$\sum \frac{k_i (f_i'' - f_i''')}{s_i} = (e^{p_0 x} - p_0 e^x + p_0 - 1) 2 N/x^3 - (p_0 e^{p_0 x} - p_0 e^x) 4 N/x^2$$

Supposons que t augmente indéfiniment, l'expression (19) tend, en vertu de ce qui précède, vers la limite :

$$(22) \quad \begin{aligned} I(x, \infty) = & - (2 \underline{0} \alpha \mu - \alpha^2 \mu) (e^{p_0 x} - p_0 e^x + p_0 - 1) 4 N/x^2 \\ & - 2 \alpha^2 \mu (e^{p_0 x} - p_0 e^x + p_0 - 1) 2 N/x^3 \\ & + 2 \alpha^2 \mu (p_0 e^{p_0 x} - p_0 e^x) 4 N/x^2 \end{aligned}$$

Cette expression permet de déterminer la limite $\varphi(x)$ de la fonction $F(x, t)$ lorsque t augmente indéfiniment ; on a en effet d'après (18), en supposant que les trois intégrales :

$$\int_0^\infty \left(\frac{\partial^2 F}{\partial x^2} - \frac{\partial^3 F}{\partial x^3} \right) \frac{x^2}{4N} dt ; \quad \int_0^\infty \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) \frac{x}{2N} dt ; \quad \int_0^\infty \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) \frac{x^2}{4N} dt$$

sont continues et tendent donc vers 0 lorsque x tend vers 0 :

$$(23) \quad \frac{d\varphi(0)}{dx} = p_0 + \lim_{x \rightarrow 0} I(x, \infty)$$

Montrons donc la continuité de ces trois intégrales pour $x = 0$.

En supposant, par exemple, $x > 0$ (sinon il suffirait d'inverser les inégalités), les inégalités (17) établies au chapitre I (p. 96) s'écrivent :

$$\left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) \left[(2 \underline{0} - 3 \alpha) x \alpha \mu + \frac{x^2}{4N} \right] \leq \frac{\partial F}{\partial t} \leq \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) \left[(2 \underline{0} - \alpha) x \alpha \mu + \frac{x^2}{4N} \right]$$

Or on a vu que $\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2}$ est une fonction ≥ 0 non décroissante de x . Soit donc $X > 0$

assez grand pour que $\left[(2 \underline{0} - 3 \alpha) X \alpha \mu + \frac{X^2}{4N} \right]$ soit > 0 ; on a alors pour $x < X$:

$$0 \leq \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) \left[(2 \underline{0} - 3 \alpha) X \alpha \mu + \frac{X^2}{4N} \right] \leq \frac{\partial F(X, t)}{\partial t}$$

ou :

$$0 \leq \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) \frac{x^2}{4N} \leq \frac{\partial F(X, t)}{\partial t} \cdot \frac{x^2}{4N(2\underline{Q} - 3\alpha)X_{\alpha\mu} + X^2}$$

Et :

$$0 \leq \int_0^\infty \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) \frac{x^2}{4N} dt \leq \frac{x^2}{4N(2\underline{Q} - 3\alpha)X_{\alpha\mu} + X^2} \int_0^\infty \frac{\partial F(X, t)}{\partial t} dt$$

$$= \frac{x^2 [\varphi(X) - e^{p_0 X}]}{4N(2\underline{Q} - 3\alpha)X_{\alpha\mu} + X^2}$$

Cette dernière expression est une fonction continue de x qui tend vers 0 lorsque x tend vers 0.

On démontrerait par un raisonnement analogue la continuité pour $x = 0$ de la deuxième intégrale ; enfin la continuité de la première intégrale résulte de celle de la troisième si l'on remarque que l'on a :

$$0 \leq \int_0^\infty \left(\frac{\partial^2 F}{\partial x^2} - \frac{\partial^3 F}{\partial x^3} \right) \frac{x^2}{4N} dt \leq \int_0^\infty \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) \frac{x^2}{4N} dt$$

La dérivée $\frac{d\varphi(0)}{dx}$ a donc bien l'expression limite (23). Or, en développant l'expression (22) suivant les puissances de x , on trouve :

$$I(x, \infty) = 2N_{\alpha\mu}p_0(1 - p_0)[2\underline{Q} - (2p_0 + 5)\alpha/3] + 0(x)$$

donc :

$$(24) \quad \frac{d\varphi(0)}{dx} = p_0 + 2N_{\alpha\mu}p_0(1 - p_0)[2\underline{Q} - (2p_0 + 5)\alpha/3]$$

Cette égalité permet de déterminer $\varphi(x)$, puisque nous connaissons la forme de cette expression ; on a en effet :

$$\varphi(x) = a(e^x - 1) + 1$$

d'où d'après (24) :

$$(25) \quad a = p_0 + 2N_{\alpha\mu}p_0(1 - p_0)\left(2\underline{Q} - \frac{2p_0 + 5}{3}\alpha\right)$$

Nous retrouvons la même fonction $\varphi(x)$ qu'au chapitre précédent (cf. p. 134).

Comparaison avec le random mating :

D'après (25) la différence $a - p_0$ entre la probabilité de fixation asymptotique du gène A et la même probabilité dans le cas du random mating, a le signe de la quantité :

$$(26) \quad e = 2 \underline{Q} - (2 p_0 + 5) \alpha / 3$$

Si l'on remarque que l'on a toujours :

$$5 \alpha / 6 \leq (2 p_0 + 5) \alpha / 6 \leq 7 \alpha / 6$$

on voit que le signe de e ne dépend pas de p_0 si :

$$\underline{Q} \leq 5 \alpha / 6 \quad \text{ou} \quad \underline{Q} \geq 7 \alpha / 6$$

Ce qui se traduit *qualitativement* en disant que, si l'un des deux génotypes homozygotes AA ou aa est nettement mieux adapté que l'autre, l'union sélective favorise sa fixation asymptotique quelle que soit la composition initiale de la population.

Par contre, si l'optimum est tel que :

$$5 \alpha / 6 < \underline{Q} < 7 \alpha / 6$$

ce qui signifie *qualitativement* que le génotype hétérozygote Aa est nettement mieux adapté que les génotypes homozygotes, le sens de l'influence de l'union sélective sur la fixation asymptotique du gène A dépend de la fréquence p_0 , c'est-à-dire de la composition initiale de la population.

Ce phénomène ne se produit pas dans le cas d'une population panmictique soumise à une sélection gamétique ordinaire : le sens de l'influence de la sélection sur la probabilité d'extinction de l'un des gènes ne dépend en effet alors que du signe du coefficient de sélection (¹).

Cas particuliers :

1° Si $p_0 = 1/2$, c'est-à-dire si dans la population initiale les gènes sont également répartis, l'union sélective favorise la fixation du gène A si $\underline{Q}$ est supérieur à α et le gène a si $\underline{Q}$ est inférieur à α . Autrement dit, la sélection favorise la fixation de l'homozygote le mieux adapté.

2° Si $p_0 = (6 \underline{Q} - 5 \alpha) / 2 \alpha$, ce qui n'est possible que dans le deuxième cas envisagé ($5 \alpha / 6 < \underline{Q} < 7 \alpha / 6$), on a :

$$e = 0$$

L'évolution de la population est alors la même que si l'on avait une population panmictique sans sélection.

(1) Cf. G. MALÉCOT, *Les processus stochastiques et la méthode des fonctions génératrices ou caractéristiques* (op. cit., p. 10).

CHAPITRE IV

DOMINANCE — MUTATIONS

Remarque sur la dominance :

Nous avons supposé (chap. I, p. 92) que le caractère primaire était sans dominance.

Les calculs et les raisonnements sont identiques à ceux que nous avons faits, si l'on suppose que ce caractère présente un certain degré de dominance caractérisé par le coefficient d tel que :

$$0 < d \leq 1$$

On trouve en effet, avec les notations déjà utilisées :

$$H_{AA} = -(2\alpha - \underline{O})^2; \quad H_{aA} = -[\alpha(1+d) - \underline{O}]^2; \quad H_{aa} = -\underline{O}^2$$

D'où l'expression analogue à celle que nous avons établie au chapitre premier (p. 92, éq. 10)

$$pT = \alpha \left\{ \begin{aligned} &P [2\underline{O}(1-d) - 4\alpha + (1+d)^2\alpha] + p(1+d) [2\underline{O} - \alpha(1+d)] \\ &+ Pp [4\underline{O}d + 4\alpha - 2\alpha(1+d)^2] + 2p^2(1+d) [(1+d)\alpha - 2\underline{O}] \end{aligned} \right\}$$

ce qui conduit pour déterminer la fonction caractéristique $F(x, t)$ de la loi de probabilité de la fréquence du gène A à une équation aux dérivées partielles du troisième ordre, semblable à l'équation (E) (chap. I, p. 95), qui s'écrit toutes réductions faites :

$$(E') \quad \frac{\partial F}{\partial t} = \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) \left\{ (1+d) [2\underline{O} - \alpha(1+d)] \alpha_{\mu} + \frac{x}{4N} \right\} x$$

$$- \left(\frac{\partial^2 F}{\partial x^2} - \frac{\partial^3 F}{\partial x^3} \right) 2x\alpha_{\mu} \left\{ \alpha [2 - (1+d)^2] + 2\underline{O}d \right\}$$

Cette équation (E') peut être étudiée comme nous avons étudié l'équation (E).

Nous nous contenterons de signaler la particularité suivante qui ne se produit pas lorsqu'il n'y a pas de dominance :

L'équation (E') n'est plus du troisième ordre si l'on a :

$$2\underline{O}d + [2 - (1+d)^2]\alpha = 0$$

ou encore :

$$(1) \quad \underline{Q}/\alpha - 1 = (d^2 - 1)/2d$$

(Ce cas est à rapprocher du cas particulier $h = (d^2 - 1)/2d$ étudié dans la première partie, chap. II, p. 17.)

(E') s'écrit alors :

$$(2) \quad \frac{\partial F}{\partial t} = \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) (x - \xi) x / 4N$$

en posant :

$$(3) \quad \xi = 4N\alpha^2\mu(1 - d^2)/d$$

L'équation parabolique (2) est du type que nous avons étudié au chapitre II (équation (3), p. 108).

Remarquons que, d'après (1), ce cas n'est possible que si

$$\underline{Q} \leq \alpha$$

c'est-à-dire si le génotype homozygote dominant est le plus mal adapté des trois génotypes.

Signalons enfin que s'il y a dominance complète ($d = 1$), les relations (1) et (3) donnent simultanément :

$$\underline{Q} = \alpha \quad \text{et} \quad \xi = 0$$

L'équation (E') s'écrit alors :

$$\frac{\partial F}{\partial t} = \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) x^2 / 4N$$

On retrouve l'équation correspondant à une population panmictique sans sélection. Ceci n'a rien d'étonnant : si les deux phénotypes sont également adaptés ($\underline{Q} = \alpha$), il n'y a plus d'union sélective.

(Ce dernier cas est à rapprocher du cas particulier $h = 0$, envisagé dans la première partie, chap. II, p. 19.)

*
* * *

Influence des mutations :

Nous avons supposé, dans ce qui précède, qu'il n'y avait pas de mutations. Faisons maintenant l'hypothèse que, pendant l'intervalle de temps Δt une proportion $u\Delta t$ des gènes a soient mutés en gènes A et qu'une proportion $v\Delta t$ des gènes A soient mutés en gènes a . Les

formules établies au chapitre premier (p. 89) sont modifiées et on obtient, en supposant négligeables $0(u^2)$, $0(v^2)$, $0(uv)$,... et en tenant compte du résultat de la première partie (chap. III, équation 2, p. 27) :

$$M_c(\Delta p) = \mu p T + u - (u + v) p$$

d'où :

$$M_c(P') = p^2 (1 + 2 \mu T - 2 u - 2 v) + 2 p u$$

$$M_c(\Delta P) = p^2 (1 + 2 \mu T - 2 u - 2 v) + 2 p u - P$$

et les formules (6) (chap. I, p. 89) deviennent :

(4)

$$A_{10} \Delta t = [\mu p T + u - (u + v) p] \Delta t$$

$$A_{01} \Delta t = [p^2 (1 + 2 \mu T - 2 u - 2 v) + 2 p u - P] \Delta t$$

ce qui conduit, par des raisonnements identiques à ceux que nous avons faits dans le cas où il n'y a pas de mutation à définir la fonction caractéristique $F(x, t)$ de la loi de probabilité de la fréquence p par l'équation aux dérivées partielles (E'') :

$$(E'') \quad \frac{\partial F}{\partial t} = x u F - x(u + v) \frac{\partial F}{\partial x} + \left(\frac{\partial F}{\partial x} - \frac{\partial^2 F}{\partial x^2} \right) [x^{\alpha} \mu (2 \underline{0} - \alpha) + \frac{x^2}{4N}] - \left(\frac{\partial^2 F}{\partial x^2} - \frac{\partial^3 F}{\partial x^3} \right) 2 x^{\alpha^2} \mu$$

avec la condition initiale :

$$F(x, 0) = e^{p_0 x}$$

La méthode que nous avons utilisée dans les chapitres précédents pour étudier l'équation (E) ne s'applique plus à l'équation (E'').

On pourrait songer à utiliser (E'') pour chercher à déterminer les moments de la loi de probabilité de p en posant :

$$F(x, t) = \sum_{i=0}^{\infty} M_i(t) x^{i/i!}$$

série uniformément convergente ainsi que ses dérivées, quel que soit x , puisque l'on a pour toute valeur de i : $0 \leq M_i \leq 1$.

On obtiendrait ainsi, entre les moments, la relation de récurrence :

$$\frac{1}{i} \frac{dM_i}{dt} = \left(u + \frac{i-1}{4N} \right) M_{i-1} - [u + v + \frac{i-1}{4N} \mu \alpha (2 \underline{0} - \alpha)] M_i + \mu \alpha (2 \underline{0} - 3 \alpha) M_{i+1} + 2 \alpha^2 \mu M_{i+2}$$

Cette récurrence (à moins de faire des hypothèses arbitraires sur les valeurs respectives des coefficients u , v , μ) ne permet pas, même pour la loi asymptotique ($dM_i/dt = 0$), de calculer les moments : en effet, l'équation obtenue pour $i = 1$ contient encore trois moments inconnus M_1 , M_2 et M_3 .

Il semble dans ce cas plus commode de rechercher directement la loi de probabilité asymptotique, qui existe et doit, d'après le théorème de Markoff, être indépendante de la fréquence initiale p_0 : le passage d'une fréquence p quelconque à une autre fréquence p' quelconque est en effet maintenant possible à chaque génération (même pour $p = 0$ ou $p = 1$) du fait des mutations.

Désignons donc par $J(p, t) dp$ la loi de probabilité de la fréquence p à l'instant t . Appliquons à l'équation (E'') la méthode indiquée par M. MALÉCOT (1) basée sur les propriétés de la transformation de Laplace.

$F(x, t)$ est la transformée de Laplace de $J(p, t)$. On a :

$$F(x, t) = \int_0^1 e^{px} J(p, t) dp$$

d'où :

$$\frac{\partial^i F(x, t)}{\partial x^i} = \int_0^1 p^i e^{px} J(p, t) dp$$

ce qui montre que $\frac{\partial^i F}{\partial x^i}$ est la transformée de Laplace de la fonction $p^i J$.

De même :

$$\frac{\partial F(x, t)}{\partial t} = \int_0^1 e^{px} \frac{\partial J(p, t)}{\partial t} dp$$

ce qui montre que $\frac{\partial F}{\partial t}$ est la transformée de Laplace de $\frac{\partial J}{\partial t}$.

En posant :

$$Y(p, t) = \frac{\partial J(p, t)}{\partial t}$$

(1) Cf. G. MALÉCOT, *Les mathématiques de l'hérédité* (op. cit., p. 47-48).

et en remarquant que par suite de :

$$Y(p, t) = \frac{\partial}{\partial t} \left[\int_0^p J(p, t) dp \right]$$

$Y(p, t)$ est nul pour $p = 0$ et pour $p = 1$, on a, en intégrant par parties :

$$\frac{\partial F(x, t)}{\partial t} = \int_0^1 e^{px} \frac{\partial Y}{\partial p} dp = - \int_0^1 x e^{px} Y(p, t) dp$$

ce qui montre que $-\frac{1}{x} \frac{\partial F}{\partial t}$ est la transformée de Laplace de $Y(p, t)$.

On en déduit enfin :

$$x \int_0^1 e^{px} Y(p, t) dp = - \frac{\partial F}{\partial t} = - \int_0^1 e^{px} \frac{\partial J}{\partial t} dp = - \int_0^1 e^{px} \frac{\partial Y}{\partial p} dp$$

ce qui montre que le produit par x de la transformée de Laplace d'une fonction de p est l'opposée de la transformée de Laplace de la dérivée de cette fonction par rapport à p .

Ces propriétés étant rappelées, écrivons l'égalité des fonctions dont les deux membres de (E'') sont respectivement les transformées de Laplace. On obtient :

$$(5) \quad - \frac{\partial}{\partial t} \left[\int_0^p J(p, t) dp \right] = J(p, t) \left\{ u - p(u + v) + p(1 - p) \alpha \mu [(2 \underline{O} - \alpha) - 2 p \alpha] \right\} - \frac{1}{4N} \frac{\partial}{\partial p} (p - p^2) J$$

équation analogue à celle qu'a donnée M. MALÉCOT ⁽¹⁾ dans le cas de la sélection ordinaire, et que nous aurions pu évidemment établir directement par la même méthode que lui, en utilisant l'expression de Δp que nous avons donnée dans la première partie (chap. III, équation 3, p. 28).

La loi de probabilité asymptotique est alors la loi stationnaire qui vérifie (3). En la désignant par $j(p)$ on a :

$$(6) \quad j(p) \left\{ u - p(u + v) + p(1 - p) \alpha \mu [(2 \underline{O} - \alpha) - 2 p \alpha] \right\} - \frac{1}{4N} \frac{d}{dp} [p(1 - p) j(p)] = 0$$

(1) Cf. G. MALÉCOT, *Les mathématiques de l'hérédité* (op. cit., équation (4), p. 48).

ou finalement :

$$(7) \quad j(p) = C p^{4Nu-1} (1-p)^{4Nv-1} e^{4N\alpha p(\alpha-2\underline{O}+p\alpha)}$$

où C est une constante définie par la condition :

$$\int_0^1 j(p) dp = 1$$

Cette loi est analogue à celle qu'a donnée M. S. WRIGHT dans le cas de la sélection ordinaire et qu'a établie M. MALÉCOT ⁽¹⁾.

L'interprétation de cette loi et sa discussion sont bien connues ; rappelons simplement que :

Si $4Nu$ et $4Nv$ sont > 1 , c'est-à-dire si les phénomènes de mutation pour un effectif donné, sont relativement importants, $j(p)$ est nulle pour $p = 0$ et $p = 1$, et est représentée par une courbe en cloche.

Si $4Nu$ et $4Nv$ sont < 1 , c'est-à-dire si les coefficients de mutation sont relativement faibles, $j(p)$ est infinie pour $p = 0$ et $p = 1$, et est représentée par une courbe en U.

La méthode indiquée par M. MALÉCOT ⁽²⁾ pour suivre au cours des générations l'évolution de la loi $J(p, t)$ définie par (5), peut évidemment s'appliquer ici ; elle montre que l'écart :

$$K(p, t) = \int_0^p J(p, t) dp - \int_0^p j(p) dp$$

entre la loi de probabilité intégrale et la loi asymptotique de probabilité intégrale tend vers 0 avec une rapidité qui est de l'ordre de grandeur de la vitesse de décroissance de l'exponentielle :

$$e^{-s_1 t}$$

s_1 étant la plus petite valeur propre de l'équation différentielle :

$$\frac{p-p^2}{4N} y'' + \left[\frac{1-2p}{4N} - u(1-p) + pv - p(1-p)\alpha\mu(2\underline{O} - \alpha - 2p\alpha) \right] y' + sy = 0$$

avec les conditions :

$$y(0) = 0 ; \quad y(1) = 0$$

(1) Cf. G. MALÉCOT, *Les mathématiques de l'hérédité*, équation (6), p. 48.

(2) Cf. *Id.*, (op. cit., p. 51 à 54).

M. MALÉCOT a montré que, dans le cas où la courbe représentative de la fonction $j(p)$ est une courbe en cloche présentant une seule dominante l_1 , et que la distribution reste, au cours des générations suffisamment concentrée autour de l_1 , $M_c(\Delta p)$ est en première approximation donnée par la fonction linéaire :

$$k(p - l) \quad (\text{cf. première partie, chap. III, p. 30})$$

où :

$$k = f'(l) \quad \text{si} \quad M_c(\Delta p) = f(p)$$

et on a :

$$l \neq l_1$$

L'équation précédente devient alors une équation de Gauss et la plus petite valeur propre relative à l'intervalle $(0, 1)$ est :

$$s_1 = -k = u + v - 6\mu\alpha^2 l^2 + 2\mu\alpha l(2\underline{0} + \alpha) - \mu\alpha(2\underline{0} - \alpha)$$

ERRATA

page 7, ligne 7 : lire ${}_1P + 2{}_1Q + {}_1R = 1 ;$

ligne 10 : Équation (3') lire

$$(3') \quad P' = {}_1p_2p(1 + \mu_1T + \mu_2T) ; 2Q' = {}_1p_2q(1 + \mu_1T + \mu_2S) + {}_2p_1q(1 + \mu_1S + \mu_2T)$$

page 53 ligne 2 : lire

$$(111) V_{III} = \Sigma \left\{ (Pi + Qj) [p''(P'i' + Q'j'') + p'(P''i'' + Q''j'')] + p'p''(Pi^2 + Qj^2) \right\} + 0 (\mu)$$

page 60 ligne 6 : lire

$$2Q(J - J_0) = 2Qn + \Sigma' [(21)m' + (22)n' + (23)r']$$

page 71 ligne 16 : lire

si l'on désigne par ρ_0 le coefficient de corrélation père-fils dans le cas du random mating

page 90 ligne 3 : lire

$$\sigma_c^2(\Delta p) = \sigma_c^2(p') \quad \text{et} \quad \sigma^2(\Delta P) = \sigma^2(P')$$

page 114 ligne 5 : lire

où β et γ sont les racines de :

page 119 Équation (28) : lire

M_i étant donné par :

$$(28) \quad \frac{\partial Z(x, t)}{\partial t} = \sum -s_i k_{ie} - s_i t f_i(x)$$

page 140 Dernière ligne : lire

Nous retrouvons la même fonction $\varphi(x)$ qu'au chapitre précédent (cf. p. 124).

page 151 Bibliographie : lire

(Publ. de l'Institut de Statistique de l'Université de Paris, 1952, vol. I, fasc. III).

BIBLIOGRAPHIE

- BLARINGHEM (L.). — *Principes et formules de l'hérédité mendélienne* (1928, Gauthier-Villars éd.).
- BARTLETT (M.-S.). — « Some evolutionary stochastic processes » (*Journal of the Royal Statistical Society*, vol. XI, n° 2, 1949).
- DAHLBERG (G.). — *Mathematical methods for population genetics* (1948, S. Karger éd., Bâle).
- FINETTI (B. DE). — « Conservazione e diffusione dei caratteri mendeliani » (*Rendic. Reale Accademia Nazionale dei Lincei*, 1927, série 6, vol. V).
- FISHER (R.-A.). — *The genetical theory of Natural Selection* (1930, Oxford).
- *Trans. of the R. S. of Edimburg*, 1918, LII, p. 399.
- GOURSAT (E.). — « Propriétés générales de l'Equation d'Euler et de Gauss » (1936, Paris, *Actualités Hermann*, n° 333).
- KENDALL (D.-G.). — « Stochastic processes and population growth » (*Journal of the Royal Statistical Soc.*, vol. XI, n° 2, 1949).
- L'HÉRITIER (Ph.). — « Génétique et Evolution » (Paris, 1935, *Actualités Hermann*, n° 158).
- LÉVY (P.). — *Calcul des probabilités* (Paris, Gauthier-Villars éd.).
- *Théorie de l'addition des variables aléatoires* (1938, Paris, Gauthier-Villars éd.).
- MALÉCOT (G.). — *Théorie mathématique de l'hérédité mendélienne généralisée* (thèse Paris, 1939).
- *Les mathématiques de l'hérédité* (1948, Paris, Masson éd.).
- *C.R. de l'Ac. des Sciences* : t. CCXV, p. 313 (1942) ; t. CCXIX, p. 379 (1944) ; t. CCXXI, p. 340 (1945) ; t. CCXXII, p. 841 (1946).
- *Annales de l'Université de Lyon (Sciences)* : 1941, p. 45 ; 1950, p. 37 ; 1951, p. 79.
- « Les processus stochastiques et la méthode des fonctions génératrices ou caractéristiques » (*Publ. de l'Institut de Statistique de l'Université de Paris*, 1952, vol. I, fasc. LII).
- MATHER (K.). — « Variation and selection of polygenic characters » (*Journal of Genetics*, 1941, vol. XLI, p. 159-193).
- ONICESCU (O.) et MIHOC (G.). — « La dépendance statistique ; chaînes et familles de chaînes discontinues » (Paris, 1937, *Actualités Hermann*, n° 503).
- PICARD (E.). — *Quelques problèmes aux limites de la théorie des équations différentielles* (Paris, Gauthier-Villars éd.).
- SIMPSON (G.-G.). — *Rythme et modalités de l'évolution* (Trad. : *Tempo and mode in evolution*, Columbia Univ. Press, N.Y., 1944), Albin-Michel éd., Paris, 1950.
- TEISSIER (G.). — *Revue Scientifique*, 1944, 82^e année, fasc. 3, p. 145.
- WRIGHT (S.). — « Evolution in Mendelian population » (1931, *Genetics*, vol. XVI, p. 97-159).

- WRIGHT (S.) — « The roles of mutation, inbreeding, crossbreeding and selection in evolution » (1932, *Proc. 6th. Internat. Congress of genetics*, 1, p. 356-366).
- « The analysis of variance and the correlations between relatives with respect to deviations from an optimum » (*Journ. of Genetics*, vol. XXX, 1935, p. 243-256).
- « Evolution in populations in approximate equilibrium » (*Journ. of Genetics*, vol. XXX, 1935, p. 257-266).
- « Statistical genetics in relation to evolution » (1939, Paris, *Actualités Hermann*, n° 802),
- « The differential equation of the distribution of gene frequencies » (*Proc. Nat. Acad. Sc.*, 1945, XXXI, p. 382-389).
- *Genetics*, 1943, vol. XXVIII, p. 114-138 ; 1946, vol. XXXI, p. 39-59.
- WRIGHT (S.) et DOBZHANSKY (Th.). — *Genetics*, 1941, vol. XXVI, p. 23-51.
-

TABLE DES MATIÈRES

AVANT-PROPOS.	1
PREMIÈRE PARTIE. — POPULATION TRÈS NOMBREUSE.	
Chapitre I. — Définition d'une loi d'union sélective.	3
Chapitre II. — Evolution d'une population nombreuse. Rôle de la dominance. . .	9
Chapitre III. — Influence des mutations sur l'évolution	27
Chapitre IV. — Passage d'une génération à la suivante. Valeurs asymptotiques des probabilités d'association des gènes	39
Chapitre V. — Corrélations relatives au caractère primaire dans une population en équilibre	57
Chapitre VI. — Fécondité et survivance différentielles. Union panmictique après sélection.. . . .	79
DEUXIÈME PARTIE. — POPULATION D'EFFECTIF LIMITÉ ET CONSTANT.	
Chapitre I. — Loi de probabilité des fréquences des gènes au cours des générations. Existence et forme de la loi asymptotique.	87
Chapitre II. — Evolution d'une population limitée mal adaptée.	105
Chapitre III. — Evolution d'une population limitée relativement bien adaptée. . .	127
Chapitre IV. — Dominance. Mutations.	143
BIBLIOGRAPHIE.	151

DEUXIÈME THÈSE

PROPOSITIONS DONNÉES PAR LA FACULTÉ

Représentation Spectrale des Opérateurs
Linéaires de l'Espace de Hilbert.

Vu et approuvé

Paris, le 29 Octobre 1953.

LE DOYEN DE LA FACULTÉ DES SCIENCES,

A. CHATELET

Vu et permis d'imprimer :

LE RECTEUR DE L'ACADÉMIE DE PARIS,

Jean SARRAILH

Imprimé sur les presses de la
Grande Imprimerie Africaine
6, Rue Carnot - DAKAR
sous le numéro 1993-5-54.
